

Aircraft Avionics

Robert G. Loewy

Georgia Institute of Technology

- I. Definitions of Avionics Components (Glossary)
- II. Aircraft Avionics Systems, General
- III. Traditional Avionics, MEP
- IV. Avionics Applications Influencing Aircraft Design; VMS
- V. Impact Of “Smart Materials”
- VI. Summary

I. DEFINITIONS OF AVIONICS COMPONENTS (GLOSSARY)

Actuator An element of a control system that will move another element, by providing a force, pressure or moment (force acting through a lever arm) in response to a command signal.

Effector A control system element that will provide the desired change in an aircrafts' behavior; e.g., aerodynamic control surface such as a “rudder,” to change heading, or a “speed brake” to reduce flight speed.

Linkage A control system component that carries useful signals, forces or moments from one location to another location. These useful signals can be analog electromagnetic or optical or digital, i.e., quantitative, and such transport can be within the aircraft or from and to points external to the aircraft. Only when forces and moments are transmitted are linkages mechanical. When digital signals are transmitted the linkages are often called “data buses.” Data buses are the conduits through which outputs are sent or inputs are received by a digital system or subsystem in order to perform its function.

Power Source Most avionics system components require power sources independent of pilot/crew; i.e., avionics systems are “active” systems. Power sources may be electrical (e.g., batteries, generators, fuel-cells) or mechanical (e.g., hydraulic pumps and reservoirs, pneumatics, etc.)

Processor A system component which may analyze (i.e., extract useful information from), combine or store signals or may model aircraft behavior for comparative purposes. Such operations may be analog or digital; when the latter, processors have much in common with computers, but usually having special, i.e., more limited functions, rather than being general-purpose.

Sensors A device that responds to some physical quantity such as pressure or temperature (or conceivably a chemical quantity such as acidity) by converting it to a useful signal.

Software The capability of digital processors and the complexity of their functions, defined above, are such that the (usually) specialized codes that command their operations are considered a separate avionics “component.” In written form such computer or processor codes may require tens of thousands or millions of

lines of instructions and their development may involve equal or greater expense than the “hardware” elements of avionics systems whose components are defined elsewhere in this glossary.

Transducer A device that takes a useful signal in one form, say electrical, and converts it to another useful form, perhaps optical. (Note that “sensors” and “actuators” are, in a more general sense “transducers,” but common usage restricts the meaning of the term as defined here.)

Transponder A component which, on receiving an ElectroMagnetic (EM) signal, often coded, will respond by sending a similar signal, usually after a known, controlled delay time.

II. AIRCRAFT AVIONICS SYSTEMS, GENERAL

The term “avionics” results from combining “aviation” with “electronics,” in recognition of the growing use and importance of the application of devices making use of electronics in aircraft design, development and operation. Aircraft avionics systems, however, make use of components which may not all be electronic, and an understanding of their functions usually requires consideration of the whole system. Figure 1 illustrates a hypothetical system for control of an aircraft about its pitch axis (i.e., pointing the “nose” of the aircraft up or down), which would “boost” the pilot’s force output in moving an aerodynamic control surface by a variable and appropriate amount, depending on the aircraft’s flight speed. In this case “the pilot’s longitudinal sidearm controller motion is converted into an electrical signal by a motion¹ sensor (Loewy, 2000). That electrical signal is converted to an optical signal by an electro-optical transducer. Fiber optic linkages carry the optical signal to a processor. After being transduced back into an electric signal, it is amplified or attenuated there according to a second signal originating from an airspeed sensor (this may be simple gain changes), so that the aircraft’s pitch response will be the same at all airspeeds (assuming this is a desirable characteristic). The signal from the processor then regulates a valve on a hydraulic actuator, which drives the aircraft’s elevator, i.e., pitch attitude control surface. Several comments may be pertinent for this illustrative example. The electromechanical input valve on the hydraulic actuator might be considered a transducer, but for our purposes it is viewed as part of the actuator. Such an assembly is often called an integrated servoactuator. Fiber optic link-

ages (fly-by-optics, FBO) are used less often than electrical linkages (fly by wire, FBW) and in FBW systems there is no need for electro optical transducers. Where fiber optics linkages are used it is usually because of their superior capabilities for carrying large quantities of information (high bandwidth) and insensitivity to ElectroMagnetic Interference (EMI), including that associated with lightning.

Once aircraft were recognized as vehicles with realizable potential for transportation, the need for a number of kinds of electronics-based equipment became apparent, based on the importance of increasing aircraft utility and safety. These functions, roughly in the chronological order in which the related avionics equipments were first adapted for use on aircraft in service use, are as follows:

1. Communication
 - a. Ground to air
 - b. Air to ground
 - c. Air to air
2. All weather, blind flying
3. Navigation
4. Limited visibility landing
5. Bad weather avoidance
6. Flight path stability augmentation
7. Improved flight handling qualities
8. Flight data recording
9. Collision avoidance
10. Formation flying (military)
11. Target acquisition (military)
12. Secure identification (military)
13. Crew/passengers comfort improvement
14. Structural load alleviation
15. Terrain avoidance
16. Noise reduction
 - a. Internal
 - b. External
17. Suppressing servo-aeroelastic instabilities
18. Performance improvement

It will be noted that this list is long, some items involve further breakdown (Items 1 and 16), and since such adaptations continue apace, any attempt at completeness is likely to soon be thwarted by new developments. For example, in-flight entertainment systems for commercial airliners are not listed, but they could be considered avionics systems. Their use is already commonplace and the services they provide are growing by leaps and bounds.

As implied by the order of functions in the list, the application of electronic devices to aircraft can be thought of as beginning with radios for communications (Items 1(a) and (b)), between aircraft crew members—pilots, copilots, navigators, flight engineers, etc.—and

¹ For illustrative purposes: side-arm controllers usually have force, rather than motion sensors.

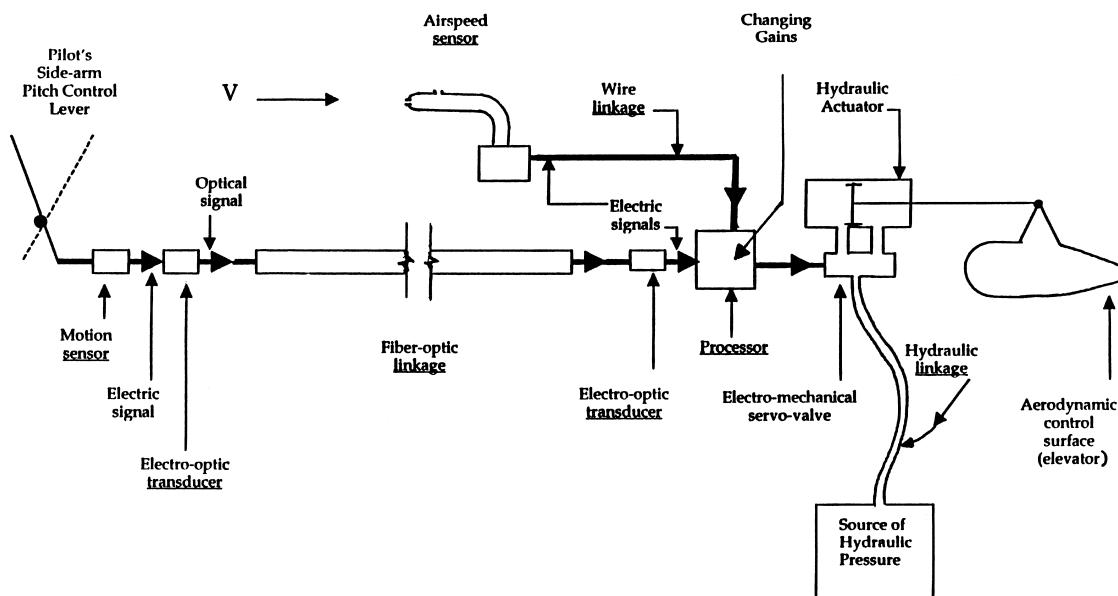


FIGURE 1 Schematic of avionics components in a Fly-By-Optics (FBO) flight control system for the pitch axis. (From Loewy, R. G. (2000). "Avionics: A 'New' senior partner in aeronautics." *AIAA J.* **37**(11), 1337–1354.)

ground crew members; among dispersed crew members of large aircraft—although this is more likely to use telephone rather than wireless technology—e.g., from the cockpit of military aircraft such as bombers, on the one hand, to the tail gunner via an “intercom,” for example, on the other; and (Item 1(c)) between flight crews of different aircraft.

Later, what can be considered radio technology, i.e., transmitters and receivers of wireless EM signals, was applied to navigation (i.e., helping the pilot know where to go) and landing aides (i.e., helping the pilot to land safely, particularly under reduced visibility conditions). In military applications of aircraft, such “assistance” by avionics for the pilot and/or other crew members was extended to acquiring and identifying targets; pointing, firing or launching weapons; countering—i.e., thwarting through so-called “electronic countermeasures”—similar systems used by the enemy; and identifying himself/herself as friendly to members of the same forces. The last is known as IFF, for “Identification, Friend or Foe.” All such functions can be performed more or less automatically to such an extent that, taken together, such systems are often referred to as “the pilot’s associate.” These kinds of avionics systems are also referred to in the military as the “Mission Equipment Package” (MEP). It is useful to think of all avionics systems which assist pilots and crew in performing their “mission,” even if it is a civilian transport moving passengers or cargo from one place to another, as MEP. Emphasizing the “added-on” nature of MEP, the aircraft in which it is installed is often referred to as the “host vehicle.”

With the advent of devices (so-called “control actuators”) capable of moving aircraft control effectors (e.g., aerodynamic surfaces) reliably and as quickly or more quickly than a human pilot, avionics systems could be used, not only to help the pilot and crew *perform their missions* but also to *fly the aircraft safely*. These automatic systems are often referred to as Vehicle Management Systems (VMS). To emphasize the highly integrated nature of VMS into the aircraft for which they are part of the control system—on an equal, flight-safety footing along with airframe structure, aerodynamic shape and propulsion systems—it is useful to think of VMS avionics as *part of* the host vehicle. As might be expected, then, VMS avionics are not “added on” but are usually considered during the design or developmental stages of the introduction into service of a new or substantially modified flight vehicle.

In a “gray area” between MEP and VMS are what, at their introduction as early as 1917, were called “automatic pilots” or “autopilots.” These systems began by using roll attitude sensors, heading sensors (e.g., magnetic compass), altitude sensors (e.g., barometric altimeters), airspeed sensors (e.g., “pitot-static” pressure tubes), etc to automatically adjust (i.e., “hold” constant) wings in a level position, aircraft direction, flight altitude and speed, respectively, by sending appropriate corrective commands, for example, to rudder and ailerons (yaw and bank effectors) and longitudinal control stick and engine throttle (speed and climb/rate of descent effectors). Such avionics systems were initially thought of as relieving pilot fatigue on long flights (particularly in an era of low cruising

speeds). With increasing avionics component capabilities, as discussed in Section IV, autopilots have developed into the much more sophisticated, Automatic Flight Control Systems (AFCS). Perhaps the most important of the enabling new capabilities is the greater responsiveness to commands on the part of actuators, i.e., their higher “bandwidth” or high frequency capabilities. Among the AFCS functions making use of these newer capabilities are included Items 4, 6, 7, and 13 through 18. The last of these could include such measures as automatic pumping of fuel from one tank to another to keep the aircraft’s center of gravity in its most favorable position on long flights.

Although avionics systems in the MEP category can, in many instances be “added on” well after the fundamental aircraft design is completed, those responsible for their integration into the host aircraft still must provide (a) space within the airframe, (b) stress-free mounting points which limit the shock and vibration transmitted to this equipment, and (c) an environment of limited maximum temperatures and EMI and acoustic fields. Such must, of course, also be provided for VMS avionics. Further, when transmitting/receiving antennae (the internal/external linkages) are involved, their locations should minimize interference with the signals to be sent or received and “cross talk” to/from other EM sources. This is particularly challenging for aircraft designed to have low radar cross sections, so-called “stealth” configurations.

Another way to categorize avionics systems in broad terms is to note that there are those for which the consequences of failures are such that a pilot can correct or compensate for them with reasonable effort, or those for which it is not reasonable to expect a pilot to do so. For our purposes, in this consideration, it does not matter whether the failure is of the type in which the system simply stops working or if it causes the system to drive to a full authority position unbidden—a so-called “hard-over” failure.

Systems for communication, navigation, or bad weather avoidance may well be important for safety of flight and hence be duplicated or provided in multiple installations of higher redundancy; but the consequences of their failure can reasonably be expected to be compensated for by a pilot if means exist to identify their improper operation. As a consequence, the duplicated systems do not usually have to operate simultaneously, but can be left in a “stand-by” mode until needed. Other systems, having to have frequency response characteristics well beyond what a pilot can do simply to operate effectively, include those for maintaining proper aircraft attitude in all weather flying, performing limited visibility landings, augmenting flight path stability, and suppressing aeroelastic instabilities. The consequences of such systems failing, therefore, will unfold much too quickly for a pilot to respond effec-

tively. The architectures of such systems should then be such as to include at least triple redundancy and a continuous comparison of performance among them in “real” time, so that a failure can be identified, in what is often called “voting,” and an automatic shut off of the system which has failed—or has even been subject to degraded performance—will take place.

III. TRADITIONAL AVIONICS, MEP

A. Communication and Navigation Systems in General

Radio technology is based, fundamentally, on the fact that an alternating electrical current (ac) in a wire will radiate EM energy into space. If the relationship between the length of the wire and the frequency of the ac, f , is such that the wire length is half a wavelength, λ , almost all the power not turned into heat in the wire will be radiated. This behavior of half wavelength wires is the basis of EM “transmitting antenna” design. A half wavelength wire which intercepts the EM radiation will also convert its energy into ac current most efficiently and is the basis for “receiving antenna” design. Note that the relationship between frequency and wavelength is given by

$$f(\text{in } H_z) = \frac{c}{\lambda} \text{ (in distance per second)},$$

where c is the velocity of propagation of EM radiation, which is that of light in a vacuum (about 300×10^6 m/s).

Sending and receiving antennae can be based (1) on the ground (terrestrial), (2) in aircraft or (3) in spacecraft. In general, the larger the antenna in terms of wavelengths of the radiation transmitted, the narrower will be the pattern of radiation. This can lead to some large airborne antennae (see Fig. 2, for example). Some antennae are designed to be omni-directional or nondirectional, i.e., they transmit EM radiation in a spherical pattern. In such a case the ratio of received to transmitted energy is equal to

$$\frac{\text{Receiver Antenna Area}}{4\pi R^2},$$

where R is the distance, or range, between transmitting and receiving antennas.

Although specific portions of the frequency spectrum (i.e., all the values of f to be used) must be allocated to prevent different systems from interfering with each other, for many years and by general agreement, radio transmission frequencies have been designated in the following “bands” (Skolnik, 1962) (Table I).

There is a marked tendency to use higher and higher frequencies, and some of the categories in Table II have also been widely used for about 50 years, but with some



FIGURE 2 Early warning E-2C aircraft. (From Skolnik, M. I. (1962). "Introduction to Radar Systems," McGraw Hill, New York.)

different applications overseas (Reference Data for Engineers (1985) ([Table II](#))).

Many factors affect the transmission of EM radiation. Some are a function of radiation frequency, others a func-

TABLE I Designated Frequency Bands for EM Radiation

Name	Abbreviation	Frequency	Wavelength
Very low frequency	VLF	3 to 30 kHz	100 to 10 km
Low frequency	LF	30 to 200 kHz	10 to 1 km
Medium frequency	MF	300 to 3000 kHz	1 km to 100 m
High frequency	HF	3 to 30 MHz	100 to 10 m
Very high frequency	VHF	30 to 300 MHz	10 to 1 m
Ultrahigh frequency	UHF	300 to 3000 MHz	1 m to 10 cm
Superhigh frequency	SHF	3 to 30 GHz	10 to 1 cm
Extremely high frequency	EHF	30 to 300 GHz	10 to 1 mm

tion of the electrical characteristics of the earth (which influence "ground waves," those propagating along the Earth's surface), atmospheric noise (such as caused by lightning), ionospheric properties (which influence "sky waves," those reflected by characteristics of the Earth's atmosphere). The influence of the Earth is, as might be expected, important for transmissions from ground stations, and is sometimes referred to as causing "site sensitivity." Very high frequencies (i.e., above 30 MHz) are mostly line-of-sight waves, and above 3 GHz, atmospheric and precipitation scattering and absorption become significant.

B. Terrestrial Based Navigation Systems

The term "avionics" usually implies equipment carried and/or functions carried out aboard aircraft. To understand some of their complexities, however, it is useful to know something of the ground-based systems with which the airborne avionics components interact. There are, in general, two kinds of ground-based navigation systems; so-called

TABLE II Letter Designation of High-Frequency EM Radiation

Letter designation	Frequency range	Letter designation	Frequency range
<i>L</i>	0.39 to 1.55 GHz	<i>X_b</i>	6.25 to 6.90 GHz
<i>L_s</i>	0.90 to 0.95 GHz	<i>K^a</i>	10.90 to 36.00 GHz
<i>S</i>	1.55 to 5.20 GHz	<i>K_u</i>	15.35 to 17.25 GHz
<i>C</i>	3.90 to 6.20 GHz	<i>K_a</i>	33.00 to 36.00 GHz
<i>X</i>	5.20 to 10.90 GHz	<i>Q</i>	36.00 to 46.00 GHz

^a Includes *K_e* band, which is centered at 13.3 GHz.

“point sources” and those that establish an EM radiation grid in space. Among point source ground systems are omnidirectional (nondirectional) beacons, which allow airborne direction finders to establish a heading direction or “bearing” to the known position of the beacon. Direction finders consist essentially of a rectangular loop antenna wired so as to send the difference in signals in the opposite, vertical sides of the loop to a “receiver.” This difference is zero when the sides of the loop are the same distance from the beacon, at which point the plane of the loop is perpendicular to the line joining the aircraft and the beacon. The antenna loop is, therefore, rotated about an axis parallel to and equidistant from the two vertical sides sensing the beacon’s signals, and its orientation must be noted when the “receiver” indicates a null reading. To minimize the aerodynamic drag on an aircraft in which a direction-finding antenna is to be mounted, two fixed antenna loops can be mounted so that their planes are at 90° to each other and the phase of their signals are compared electrically from one to the other to achieve the same effect as mechanical rotation. Direction finding systems can, alternatively, have the beacon placed in the vehicle and the rotating loop antenna and receiver at the ground station.

The nearly constant speed of EM radiation has led to its use to measure distance. Although other means of radio ranging (i.e., means to measure distance) exist, perhaps the simplest in concept is known as DME, for Distance Measuring Equipment. This system is internationally standardized. Its operation is depicted in Fig. 3 (from Kayton and Fried, 1997). A transmitter-receiver on board the aircraft, known as an “interrogator,” sends a pair of very short EM pulses (3.5 μ s long and 12 μ s apart), repeated from 5 to 150 times per second. A transponder at a fixed, known ground station, on receiving these pulses, retransmits them after a 50 μ s delay. The avionics component on the aircraft automatically determines the difference between sending and receiving times (very short compared to the period of the highest repetitive rate) subtracts the transponder delay and shows the distance from the ground station on a control panel display.

Bearing, as provided by direction finders, and distance, as provided by a DME, from the same ground station, allows the calculation of position relative to that ground station. In geometric terms, by establishing range and bearing, ground “point sources” allow an aircraft to place itself on the space curve intersection of a sphere (from range information) and a semi-infinite vertical plane which has one edge at the fixed ground station (from bearing information). If barometrically determined altitude information is added, the position of the aircraft will be known.

Another kind of ground-based point source is intended to provide aircraft occasional, positive and absolute location information, often known as a “fix.” These EM radiation transmitters are known as “marker beacons” and they send a narrow, fan-like pattern vertically at fixed points along the nation’s airways, with the pattern’s maximum width aligned with the center-line of the airway on which they are located. Receivers in the aircraft provide the pilot with information as to which beacon has been or is being traversed.

The VHF (Very High Frequency) band listed in Table I is used for voice communications to, from, and among aircraft. By combining communications and navigation functions in the VHF band, some avionics components can be made to do double duty. The success of this scheme has resulted in what is known as VOR (for VHF Omnidirectional Range) and its adoption as an international standard. In this system, the ground station radiates two signals: one is omni-directional radiation whose carrier VHF frequency is modulated at 30 Hz; the second is a cardioid (heart-shaped) pattern in the horizontal plane that rotates at 30 rps. The airborne receiver experiences both transmitted waves as 30 Hz signals and the phase angle between them, as related to the rotation angle of the cardioid pattern, determines the bearing of the VOR beacon

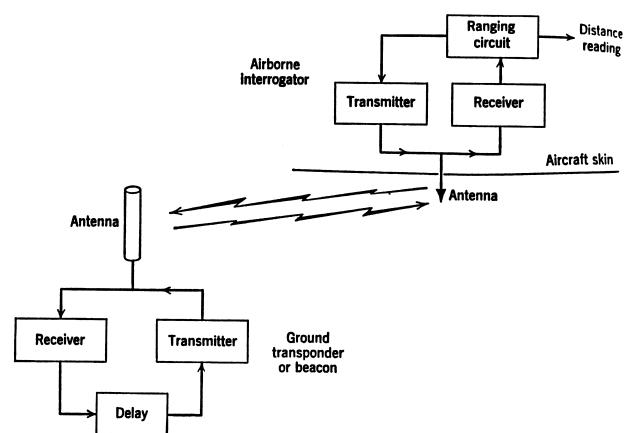


FIGURE 3 DME operation. (From Kayton, M., and Fried, W. R. (1997). “Avionics Navigation Systems,” Wiley, 2nd ed., New York.)

from the aircraft. There is also a Doppler version of VOR (see Section III.E) in which a 9960-Hz carrier frequency is frequency modulated by the (simulated) rotation of a large diameter (480 wavelengths or 44-ft-diameter) antenna² so as to be varied by ± 480 Hz at 30 Hz. The same airborne equipment can be used to sense phase, hence bearing to the Doppler VOR beacon, as with ordinary VOR, but with less site sensitivity and greater accuracy. Maximum bearing errors at 20-mile distance with standard VOR are about 3° and with Doppler VOR, about 0.5° .

The military uses "point source" ground stations which combine systems for determining both bearing and distance measurement. These systems are known as Tacan (Tactical Air Navigation). The distance measuring function, i.e., range determination, is accomplished using the same pulse and frequency configurations as standard DME. The Tacan omni-bearing operation, however, (a) uses frequencies from 960 to 1215 MHz (almost 10× higher than VOR) so that smaller antennas can be used; (b) employs a multi-lobe radiation principle which improves bearing accuracy; and (c) enjoys equipment economics as a result of using the same radio frequencies for range and bearing determination.

The so-called "hyperbolic" systems, such as Loran, Omega and Decca, provide an alternative means of position determination. These systems, rather than using "point sources," consist of groups of transmitting stations thought of as forming "chains." A chain consists of at least three stations, of which one is a master transmitter and the other two are secondary transmitters. Each station in a chain transmits EM pulses which are grouped closely in time and repeated at a certain rate. The interval between the repeated transmissions of these groups of pulses is known as the Group Repetition Interval (GRI), and it identifies a particular chain. The number of pulses in a group, the interval between them, the envelope which defines pulse shape, as well as GRI, establish the transmitted signal format, and it identifies each station in a chain. Since the positions of the stations are known, as are the timing of signals transmitted from master and secondary stations, the difference of the Time Of their signals' Arrival (TOA) at an aircraft informs the aircraft that it must be somewhere on a space curve, which happens to be a hyperbola, in a horizontal plane determined by barometric altitude. A series of these TOA's, then, establishes a series of hyperbolas, as shown in Fig. 4 (from Kayton and Fried, 1997). It follows that TOA's from that master and another secondary station establishes a second series of hyperbola. The two specific TOA's informs the aircraft as to which two hyperbolas it must be on; their intersections (plus barometric altitude) establish the aircraft's po-

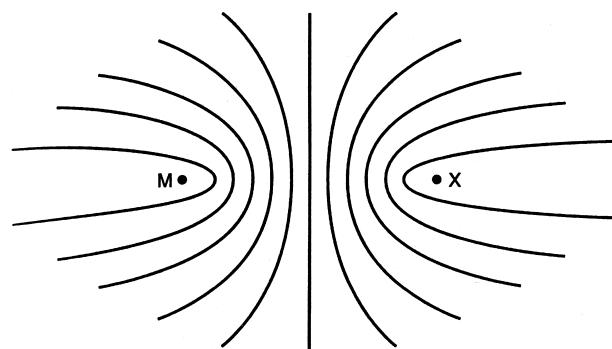


FIGURE 4 Hyperbolic lines of constant TD for a typical master-secondary pair. (From Kayton, M., and Fried, W. R. (1997). "Avionics Navigation Systems," 2nd ed., Wiley, New York.)

sition in space, as shown in Fig. 5 (Kayton and Fried, 1997).

Since LORAN-C is affected by sky waves and uses ground waves, sophisticated corrections must be made to achieve maximum accuracy, and the usual ranges, a function of transmitter power, are measured in hundreds of miles. Accuracy of about 1/2 km is achieved roughly 95% of the time, when differential techniques (see, for example, Section III.C) are used, adding redundant station pairs.

In the Omega system (which has been shut down for several years), transmitters emitted continuous waves, rather than pulses, and hyperbolic lines of position were established by phase differences in signals received from a master/secondary station pair. Because one phase difference between two continuous wave signals defined a series of hyperbolas, multiple frequencies were used to eliminate the ambiguity of which hyperbola was the pertinent one.

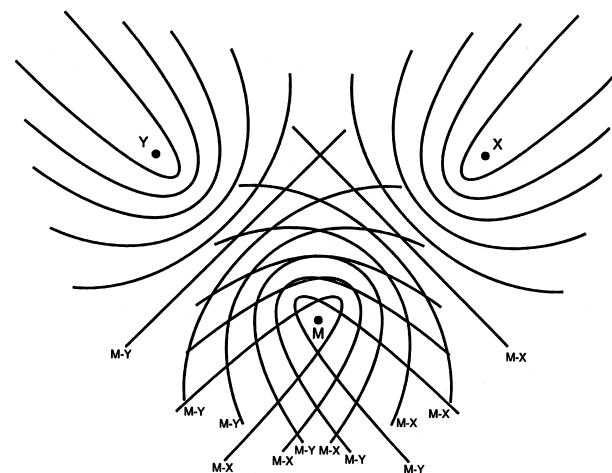


FIGURE 5 Hyperbolic lines of constant TD for a typical triad. (From Kayton, and Fried, W. R. (1997). "Avionics Navigation Systems," 2nd ed., Wiley, New York.)

² Actually a ring of individual EM transmitting elements.

Five frequencies were used at each station; four common and one station-unique. The frequency bands being VLF (see Table I), these signals were propagated with low attenuation between the Earth's surface and a particular layer of the ionosphere, to great ranges. In fact, only 8 transmitting stations worldwide constituted an Omega system with accuracy within about 4 nm, 95% of the time. When 30 such stations were employed, using differential techniques among redundant pairs, position errors were diminished to about 2 km, 95% of the time, within 1000 km of the monitor station.

Decca is also a system based on phase differences, but one that uses low frequency carrier waves between 70 and 130 kHz. Two station pairs are typically 110 km apart and the range of coverage is typically 320 km.

With the advent of digital technology, both the accuracy and flexibility of the earlier systems' information processing have increased. Further, the components', particularly the airborne components', size and weight have been reduced. Since many aircraft radio communication and navigation systems have shared both the same parts of the EM frequency spectrum and a common technology, both ground-based and airborne components of terrestrial systems which provide digital communication and navigation functions have been developed as integrated systems. That is, the same EM carrier waveforms are used to carry both functions.

Two basic types of terrestrial integrated communication-navigation systems, centralized and decentralized, are in widespread use by the military. Operation of the former is dependent on a central site, from which all users determine their positions on an absolute basis. Such an arrangement facilitates the control of many users, more or less simultaneously, although users ordinarily receive information on their positions automatically on the basis of periodic "requests" from the ground-based central site or node. Users in these "Position Location and Reporting Systems" (PLRS) are "cooperating" users; i.e., they are equipped with Radio Sets" (RS) equipped with accurate "clocks" and send a signal, individually identifiable, to three or more ground stations (MS's for Master Stations). The MS's also have very accurate "clocks" and comparisons of their timing signals are made with those of the RS. Knowing "clock" signal differences and signal TOA information at two MS's, places the user (the RS) on the space curve intersection of two (imaginary) spheres. Altitude information, based on barometer data, establishes the position of the RS at one of two points on this space curve, and TOA data from a third MS eliminates this uncertainty.

Operation of the second, decentralized type of system is such that each user determines its own coordinates relative to other users' positions. Since it is independent of central sites, decentralized systems are often called "node-

less." Decentralized systems clearly require the airborne components to be both receivers and transmitters.

C. Satellite-Based Navigation Systems

Relatively soon after the successful orbiting of man-made satellites about the Earth, attempts began with the objective of replacing or supplementing terrestrial radio navigation systems through the use of earth satellites. The major advantages provided are those of (1) coverage, since the lines of sight, with enough satellites in the proper orbits can be made to reach all points on earth; and (2) very stable operation. Their transmission frequencies (L band navigation signals from the satellites and S band telemetry downlink to and up-link from the ground station) are also such as to make them all-weather systems. Two major satellite radio navigation systems are included in an internationally recognized Global Navigation Satellite System (GNSS); they are the U.S. Department of Defense's NAVSTAR Global Positioning System (GPS) and the Russian Federation's Global Orbiting Navigation Satellite System (GLONASS). Both systems have three elements: (1) a constellation of earth-orbiting satellites (each has 24, as of this writing), (2) ground stations; and (3) receiver/processor units in the user aircraft. The satellites transmit EM signals which the ground station uses to track them and from which the user aircraft determines its position relative to the satellites. Very accurate atomic clocks aboard the satellites are the heart of GPS. TOA processing in the ground stations allow simultaneous ranging from multiple locations, and since the ground stations' positions are known—also allows determination of the satellite's location, velocity, and predicted orbital positions. The satellite's orbital position information is sent from the ground stations to the satellites, which transmit it to the user aircraft's receiver processor, together with timing signals. The user aircraft's processor uses TOA data to establish its position relative to (at least three) of the GPS satellites which, combined with the transmitted data on *their* positions, allows the user aircraft's position to be known. An obvious advantage from the avionics viewpoint of GPS type navigation systems is that the equipment in the user aircraft can be "passive" to the extent that user aircraft need not transmit EM signals.

A variety of corrections are required in GPS or GLONASS to achieve the position accuracy desired; such include those compensating for clock errors, the rotation of the earth, ionospheric and tropospheric refraction, etc. "Differential" principles can be used to eliminate errors common both to the user and a reference ground station. In "Differential Global Positioning Systems" (DGPS), a "reference ground station" receives the same navigation signals as the user aircraft, but since its position is known,

all the errors in its calculated position can be determined. These become error corrections when the reference ground station transmits them to the user aircraft. Through the use of DGPS, position errors can be reduced to within 1 to 10 m, depending on the user's distance from the reference station.

GPS receivers can be quite small; one military version is known as the MAGR (for Miniature Airborne GPS Receiver) and the entire component, except for its antenna, is contained inside another avionics assembly; e.g., an inertial navigation system (See Section III.D, below). The antenna itself, in a U.S. Navy system, is contained in a circular housing with a diameter of less than 125 mm, has a height (thickness) of about 40 mm, and weighs about 2 n (mass of about 0.1 kg).

D. Inertial Navigation Systems

The basis of inertial navigation is "dead reckoning" (see Section III.E, below), using accelerometers mounted on the aircraft to measure accelerations and integrating their signal outputs over time, first to obtain velocities and then a second time to determine position. Inertial Navigation Systems (INS) are self-contained, requiring no cooperating ground stations or satellites sending EM signals to the user aircraft; thus, they are not subject to interferences by an enemy or the weather. Since the processing of the fundamental sensor output is an integration over time, however, errors grow with time, and, if the orientation of the accelerometers is not accurately known, aircraft attitude changes—as a result of atmospheric disturbances or deliberate maneuvers—will contaminate acceleration signals with changing gravity components. Corrections for these effects are accomplished in so-called "strap-down" inertial systems in which gyroscopes are added to sense angular motions. These "strap-down" inertial systems have become practical with the advent of Ring Laser Gyros (RLG's) and Fiber Optic Gyros (FOG's). These devices correct for changes in acceleration direction electrically, so that the linear, horizontal velocity and position predictions are as they should be for the purposes of navigation.

When inertial systems are activated, they must be "aligned," to set the aircraft's initial position and velocity properly and to orient its axes relative to the Earth; this process is known as "gyrocompassing." The Earth's rotation, of course, imposes a centripetal acceleration whose magnitude and direction (with respect to the "vertical," i.e., the local normal to the earth's mean surface) varies with geographical position. This and other such small errors grow sufficiently with time as to make "hybrid" systems, such as those which "update" inertial systems periodically using GPS data, for example, in a process known as "aiding"

worth their additional complexity. INS is widely used in the military and on large civil passenger aircraft.

As principal components of inertial systems, accelerometers and gyroscopes have been subject to intense development efforts to improve their accuracy and eliminate responsiveness to influences which contaminate their outputs. Design of accelerometers for inertial navigation systems are most often based on one of three concepts. The first is that of a pendulum on "flexures"—beams with very low stiffness in one direction, but stiff in the other two, perpendicular directions—and electrically restrained to a zero deflection at zero or reference acceleration. This provides for "rebalancing" to ensure that response to one acceleration will not change the direction of sensing for the next. The second makes use of very small, micromachined silicon masses mounted on springs that are "soft" in one direction, stiff in the other two, also electrostatically nulled; and the third employs vibrating beams whose stiffness is so low in the direction of vibration that tensile force variations along the beam length cause changes in the frequency of vibration, thus indicating acceleration along the beam length.

Many types of gyroscopes are used in aircraft applications for either indicating or providing signals in automatic systems which provide control of aircraft attitude angle or angular rates. The earlier forms used a spinning wheel mounted in a gimbal (so as to be free to rotate about an axis perpendicular to its spin axis) and "floated" at neutral buoyancy. Angular motion of the gimbal axis in a plane containing the spin axis would then cause precession about the gimbal axis, which would indicate the gimbal axis' angular rate. If this response were to be available again for later motions, this precession angle would have to be "reset," and such would be done by magnetic torquers, according to a "rebalance algorithm." These and other gyroscopes were developed further, including such refinements as two perpendicular gimbals, electrostatic suspension, etc. The less expensive, less maintenance prone versions with drift rates of about 0.1 deg/h are still useful, for example, in tactical missiles, but are too inaccurate for long-range navigation.

Although modern optical angular motion sensors are still called "gyroscopes," they function on other than Newtonian mechanical, i.e., "inertial," principles. Because of their accuracy, dynamic range, linearity, maintenance-free nature, and reliability, RLG's and FOG's are now used in INS's for almost all commercial and military aircraft. One of two optical principles are used in these devices, but in either application, two laser beams propagate in what is essentially the same closed, planar path; one clockwise, the other counterclockwise. If the device containing the paths rotates about an axis perpendicular to the plane of those paths, the Sagnac Effect (1967), which results from

the fact that light waves are unmoved by motion along the light path of the medium in which they're transmitted—makes the light's path in the direction of device rotation appear to be longer and the light's path in the opposite direction appear to be shorter. The RLG makes use of a “resonator” principle, the FOG can use that principle or interferometry.

Because the front part of a laser light beam is coherent (i.e., all components are in phase) the interference between two beams propagating in opposite directions in an optical resonator forces a standing wave within the optical cavity. When this type of RLG's housing rotates about the circular path's centerline, then, the nodes and/or antinodes of the standing wave, which are fixed in space, can be “counted” and interpreted as angles in the azimuthal direction around the circular path. The light sensor of an FOG using the interferometry principle experiences phase differences where the two, counter-rotating laser light beams emitted simultaneously are recombined, since one's path is longer and the other's shorter, depending on the sense and magnitude of the angular rotation of the device about the path's centerline. The positions of the lines of interference can then be interpreted as a measure of rotation angle. Most optical gyros used in INS's, as of this writing, are of the interferometer type; employing light paths of between 10 to 40 cm in length; weighing between 5 and 20 n (mass between 0.5–2.0 Kg) per axis; and having root-mean-square accuracies of about 0.05°.

The typical INS, using these components, then ([Kayton and Fried, 1997](#)), requires about 8000–16,000 cc in volume, 30–150 W of power, weighs approximately 85–130 n (mass between 9–14 Kg) and has a velocity accuracy of about 0.75 m/s (rms) and navigational accuracy of 1.5 km/h. These modern airborne systems are relatively expensive (\$50,000 to \$120,000).

E. Doppler Radar and Dead Reckoning Systems

“Dead reckoning” is an old maritime term used to describe navigating (itself a maritime term) by using known initial position, the vehicle's velocity vector (speed and direction), and how long that velocity has been maintained, to determine the vehicle's new position. If velocity is measured, say, relative to the surface of a body of water, it is clear that positions determined by dead reckoning will be in error by the existence of currents in that water. For ships whose speed is not great relative to currents, this is important; for fast flying aircraft it is much less so. Use of Doppler radars to measure relative speed in modern dead reckoning systems, however, has some significant advantages; for example, like INS, they are self contained, needing no terrestrial or satellite cooperative station; their

transmitter power requirements are small; and they work very well for low vehicle velocities.

As to the principle on which Doppler radars are based, consider that wave motion emanating from a source moving with respect to the receiver is sensed as having a changed frequency; the magnitude of the change depending on the relative velocity, higher if the source and receiver are moving closer, lower if they are moving farther apart. This so-called “Doppler effect” is experienced almost every day acoustically, for example, if a fast-moving auto or train passes with its horn or whistle blowing. In a directly analogous way, the frequency of a radar signal return shifts if the transmitter and reflecting surface have relative velocity along the line of EM transmission. This provides a means, using reflection returns, to determine the speed of an aircraft relative to the ground or water over which it is flying. Doppler radars, mounted on an aircraft, use microwave frequencies in an internationally authorized band, between 13.25 and 13.4 GHz. This provides narrow beams of EM radiation, which can be pointed at the ground at relatively steep angles. The last has the additional benefit of reducing the probability of detection in military applications.

For Doppler navigation, at least three radar beams are needed to determine three components of velocity relative to the earth's surface, and three aircraft attitude measurements in three perpendicular planes are needed to resolve the Doppler radar measurements into components in an earth-related, geodetic coordinate system, as needed for dead reckoning navigation ([Fig. 6](#)). If the three Doppler radar beams are arranged as shown in [Fig. 7](#), and a difference taken of the returns from signals A and B, the Doppler shifts of the lateral components will cancel, whereas the longitudinal components, being of opposite sign, will be added. This arrangement, known as a “Janus” system (after the Roman God who could see both backward and forward), increases system accuracy. For the usual beam angles to the horizontal of about 70°, a Janus system will

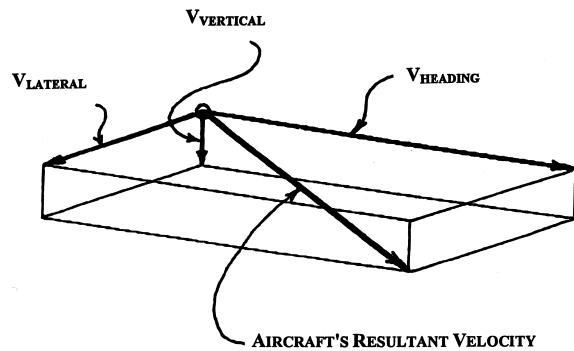


FIGURE 6 Resolution of aircraft velocity into navigable components.

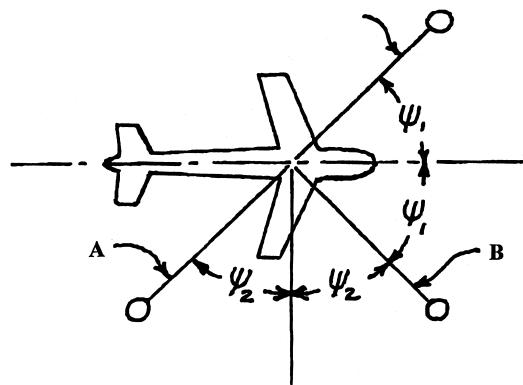


FIGURE 7 Lamda arrangement of three Doppler radar beams.

have an error in horizontal velocity of only about 0.015% per degree of error in the aircraft's pitch attitude. Without the Janus arrangement the same metric would be about 5%.

F. Instrument Landing Systems

What is known as the Instrument Landing System (ILS) consists of (1) "localizer" transmitters, located at the centerline of and off the ends of runways, which provide lateral guidance to aircraft approaching to land; (2) "glide slope" transmitters located beside runways near the end of the runway over which the aircraft first passes in landing (the "threshold"), which provide vertical guidance; and (3) marker beacons reporting progress along the glide-path to the pilot of the landing aircraft.

All three, localizer, glide slope, and marker beacon transmitters, radiate continuous wave EM energy at radio frequencies. Their radiation patterns in space and in specific frequency channels provide signals to an aircraft receiver indicating deviations from the desired height as a function of range from the end of, and lateral displacement from the centerline of, a particular runway. The receiver, then, displays information to the pilot that only is "nulled" when the aircraft is "on course" to landing, and these information signals grow with the level of deviation to either side or above/below the proper course. In "auto land" systems, such deviation signals are "hard-wired" to the AFCS (see Section IV). In the ILS airborne equipment, a Morse-Code identification signal is received audibly in the cockpit on the localizer band, and a voice transmission from the airport's control tower may also be provided. Signal standards for ILS are established internationally, and about 1500 ILS's are operational at airports throughout the world.

G. Collision Avoidance Systems

Although ground control of aircraft flight plans and flight paths in real time have as a prime objective eliminating

the possibility of mid-air collisions, controlled flight into terrain, or collisions on airport taxiways or runways, vehicle borne avionics equipment plays important roles in these functions. Installation of airborne, mid-air collision systems provides protection against such calamities independent of ground control and in addition to the nation's Air Traffic Control (ATC) system. The Traffic Alert and Collision Avoidance System (TCAS) uses a scanning radar transmitter in one aircraft to trigger the response of a transponder in any aircraft so equipped within its range. In the version of TCAS most used, TCAS II, the distance between two aircraft and their altitude separation are calculated, based on the transponder signal returns, and the crew is alerted about 40 sec before the closest Point of Approach (CPA) is to be reached, if the separation is predicted to be small. This alert is known as a "traffic advisory" and displays the range, bearing and altitude of the aircraft posing collision danger. If the danger continues, about 25 sec before CPA, a "Resolution Advisory" (RA) appears showing the climb or descent maneuver recommended to increase the miss distance. Since both TCAS-equipped aircraft must be properly advised as to how to change their flight paths, TCAS II has an air-to-air data link communicating between the two aircraft, to coordinate RA's. A version known as TCAS I does all of this except displaying RA's. All air carrier aircraft operating in U.S. airspace with 10 to 30 passenger seats must have TCAS I, all with more than 30 seats must have TCAS II.

As air travel departure and arrivals increase and, with them, airport congestion, the danger of collision between aircraft on the ground also increases. Ground-based systems to aid the regulation of ground movement of aircraft include surface movement radars and taxiway lights modulated to indicate specific taxi routes. These don't, at this writing, require avionics equipment on the aircraft. There are, however, aircraft systems that use transponders to allow ground-based "interrogators," located with taxiway lights, to derive identification and location information and relay it to tower controllers.

Commercial, in fact all civil, aviation must keep a safe altitude above terrain in all flight modes other than take-off and landing. Military aircraft, however, must often approach the ground for weapon delivery, precise reconnaissance, or—for extended flight times—to avoid detection or defensive weapons. Those with the last of these requirements are usually equipped with "terrain following" equipment. These are automatic systems having Forward Looking radars and Infra-Red (FLIR) sensors and radar altimeters (see Section III.I, below). The first two of these measure the range and angle from the horizontal of the terrain before the aircraft. Flight path control commands, based on a computed terrain profile in a vertical plane, based on the forward-looking radar returns with

pilot monitoring, directly control the pitch attitude, hence “angle of attack” and lift of the aircraft, so as to maintain a desired height above the terrain. The radar altimeter checks the altitude prediction, and the FLIR provides back-up data to ensure that the radar commands flight over such obstructions as power lines. In civil aviation, Controlled Flight Into Terrain (CFIT) has become a safety issue of increasing concern, particularly in mountainous regions and under conditions of reduced visibility. Appropriate avionics therefore, may soon be appearing on all aircraft above a certain size in terrain avoidance applications.

H. Weather Radar Systems

Airborne weather radar systems make use of the reflectivity of clouds, precipitation, dust particles at low altitudes and ice crystals at high altitude. Their intent is to determine the position of air having the kinds of motion that can make flight dangerous or uncomfortable. The magnitude of reflections, compared with stored models, allows precipitation rates and whether it is rain, snow or hail, to be determined. Positions are determined by the elevation and scanning (heading) angles of return signals and range by “gating,” i.e., “enabling” the receiver only at the specific times when a signal reflected from that range would be received. Further, the wind speeds and turbulence intensities are measured by Doppler effects, usually by processing the return timing of pairs of pulses.

Weather radars and multiple use radars with weather defining functions are usually mounted in the nose of the aircraft (see Fig. 8, from Kayton and Fried (1997)) carrier wavelengths in either C or X band are used, and the weather displays usually show weather formations within about 100 km ahead of the aircraft. Warnings can be given to the pilot if stored levels regarded as the maximum al-

lowable for safe operations are exceeded for wind speeds or turbulence; 5 m/s, for example, is often considered such a limit for turbulence. The atmospheric phenomena associated with thunderstorms and “microbursts” and known as “wind shear” is susceptible to detection using Doppler processing avionics. This extremely dangerous condition is associated with a column of air with high downward vertical velocity flows which, on contact with the ground must flow horizontally (i.e., radially) outward in all directions. An unsuspecting aircraft flying directly toward the vertical column, near the ground (as in a landing approach), is flying into a head wind until it passes the center of the column, at which time it is abruptly subject to a tailwind. The associated loss of lift may not be restored before disastrous contact with the ground. Weather radar to sense these potentially calamitous conditions are installed at a few airports. Plans are underway to install weather radars specifically capable of detecting wind shears on all commercial aircraft with detection range capability sufficient to allow this atmospheric phenomenon to be evaded.

I. Radar Altimeters

For certain specialized functions, down-looking radars are used to measure “tape-line” altitude; i.e., distance above the ground immediately below the aircraft. Low flying military aircraft may use radar altimeters for correlating terrain contours with stored map-matching navigation information. In civil aircraft operation, radar altimeters can be used to assist automatic landing systems. Although other techniques have been used, in one currently used radar altimeter, a continuous wave is generated whose frequency is modulated with a period much longer than the time required for the ground-reflected signal to return to the aircraft. Comparison of generated frequency with returned

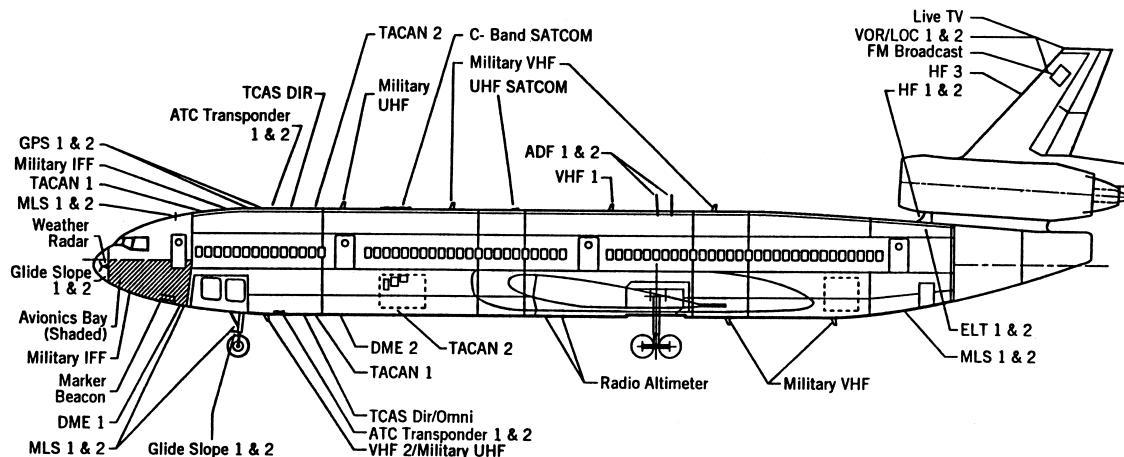


FIGURE 8 Avionics placement on multi-purpose transport (From Kayton, M., and Fried, W. R. (1997). “Avionics Navigation Systems,” 2nd ed., Wiley, New York.)

frequency for the known modulation schedule, then, is a measure of elapsed time, hence the distance above the terrain. Antennae beam width, for this function, must be wide enough to provide a return from a flat earth when the aircraft pitches and rolls over the normal range. Radar altimeters have the 4.2 to 4.4 GHz frequency band assigned to them, allowing use of two small, microstrip antenna's, one to transmit, one to receive, producing beam widths of between 40 and 50°. If the separation between this pair is at least about 0.8 m, the transmitted/received signals' mutual interference will be reduced enough that, even after the loss of signal intensity experienced in ground reflection, useable information will still remain in the received signal. Radar altimeters are usually only used for tape-line altitudes of less than 1500 m; current continuous wave versions require less than 1 W of power and have accuracies of about 0.5 m below 30 m and about 2% above 30 m.

IV. AVIONICS APPLICATIONS INFLUENCING AIRCRAFT DESIGN; VMS

A. Flight Path Stability; AFCS

Means to provide equilibrium, stability, and control for aircraft are obviously essential. Yet devices and arrangements of aerodynamic surfaces to carry out these functions are usually detrimental to such aircraft performance as how far the aircraft can carry a given payload, its maximum speed, etc. Tail surfaces, for example, add weight and aerodynamic drag, and having the aircraft CG forward of the wing's center of lift as required for positive natural, i.e., unaugmented, stability in pitch, requires a download on the tail for equilibrium. The present-day capabilities of avionics, with actuators capable of high-frequency response to control inputs, allow designers to deliberately take advantage of the often superior range-payload, high-speed, and high-maneuverability performance of configurations whose unaugmented flying qualities are grossly unsatisfactory, by incorporating an avionics-based Stability Augmentation System (SAS) to compensate for unstable flight path characteristics. [Figure 9](#) (from [Loewy \(2000\)](#)) illustrates one of the advantages of Reduced Static Stability (RSS) in pitch that was exploited in the basic design of the F-16 Falcon jet fighter. By integrating so-called RSS into the basic aircraft design, equilibrium of pitching increments was achieved with the horizontal tail generating lift rather than download, so that the lift that the wing must produce to carry the weight of the aircraft is reduced. This means a smaller wing with less structural weight and drag can be incorporated into the overall design. Further, for equal offsets of the wing's lift from the aircraft CG, either fore or aft, the tail lift (with RSS) can be less than the

tail download (with positive static margin) so that the tail's drag, i.e., "trim drag," will be less. The cumulative effect of these beneficial design changes, resulting from allowing an avionics SAS to compensate for RSS, is to reduce the fuel required for the same range-payload combination and a higher net thrust-to-weight ratio. The latter is particularly important for fighter aircraft because it improves maneuverability.

If the unaugmented flying qualities of an aircraft are sufficiently poor to constitute a risk to flight safety, however, the reliability of the SAS must be extremely high, e.g., approaching that of primary structure. Further, provisions must be made to assure the kind of controllability that avoids hazardous aircraft attitudes and/or angular rates even in the unlikely event of "hard-over," i.e., abrupt, full control travel, system failures. In the case of the F-16, this resulted in a quadruply redundant SAS. Implications for the aircraft's design went beyond the fundamental performance and maneuverability characteristics mentioned before. As stated in ([Droste and Walker \(1990\)](#), "Since the pilot would not be able to control the aircraft in the pitch axis without the electrical system, there was no justification to retain a mechanical pitch system. The pilots' command could now be electrically combined with the stability system with no penalty. Removal of mechanical connection between pilot and the control surfaces was the logical result." That is, making the SAS highly reliable and fail-safe allowed the use of an FBW flight control system with no mechanical control system to back-up the electrical system. FBW control systems also facilitate means to enhance the way an aircraft responds to a pilot's command. These aircraft avionics systems are referred to as Control Augmentation Systems (CAS).

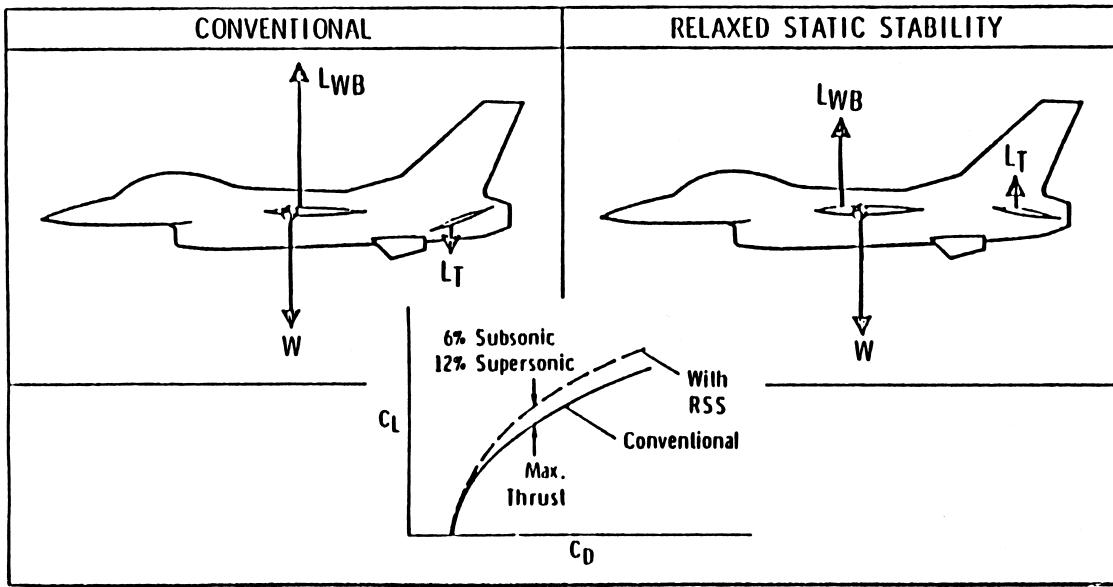
It should be clear that avionics systems capable of SAS and CAS functions can also perform "autopilot" functions as mentioned in Section II, above. When all these functions are integrated into one system, it is commonly referred to as an Automatic Flight Control System (AFCS).

B. Load Alleviation

Another example of avionics systems usage is in maneuver load-limiting. The prototype of the F-16, the YF-16, was designed for a limit normal load factor (n) of 9 times gravity, or "9g", throughout most of its flight envelope, but as low as 6.5g in some critical areas. To prevent pilots from exceeding these limits, pilot commands were attenuated by avionics-based controls as a function of Mach number (forward speed divided by the speed of sound), altitude, and Angle Of Attack (AOA). This system also limited AOA in an absolute sense, i.e., independent of Mach number and altitude. The associated automatic AOA (or n)—limiting schedule is shown in [Fig. 10](#) (from [Loewy](#)

RELAXED STATIC STABILITY

- The Benefits of Relaxed (Reduced) Static Stability are Integrated into the Basic Aircraft Design



- Less H.T. Deflection (T.E. Up) → {

- More Aerodynamic Lift
- Less Trim Drag

(WHICH RESULTS IN:

1. Reduced Structural Weight (Smaller Wing)
2. Reduced Fuel for Same Range (or more range, same fuel))

FIGURE 9 Relaxed static stability. (From Loewy, R. G. (2000). "Avionics: A 'New' senior partner in aeronautics." *AIAA J.* 37(11), 1337–1354.)

(2000)). The production version of the F-16 was designed with a 9g limit normal load factor throughout the operational envelope, eliminating the need for this particular avionics system, but roll rate (angular velocity about the aircraft's longitudinal axis) limiting was retained with variable limits as a function of AOA as a means of ensuring good handling qualities at the extreme maneuver attitudes at which strong dynamic coupling between angular motion about two inertial axes exists.

A second example of a load alleviation avionics system is provided by the arrangement used to extend the wing structure's fatigue life as an interim measure during the development of the Lockheed C-5A military transport airplane. The operational concept on which this automatic system was based followed from two physical laws, one aerodynamic, the other in mechanics of materials. First, a wing will have minimum drag induced by lift (induced drag) if the spanwise lift distribution approaches an elliptical distribution. This will maximize range-payload perfor-

mance, assuming that proper account is taken of the structural, i.e., weight consequences, attendant on the bending and shear loading associated with such a spanwise distribution of lift. The second physical law is often expressed in so-called "Goodman Diagram" form: (McClintock and Argon (1966), for example), which shows that the lower the *steady* stresses in a material, the higher are the allowable *alternating* stresses for which the material will have indefinite life; the so-called endurance limit. Because a major source of alternating stresses on airplane wings is "gusts" (i.e., atmospheric turbulence) in cruise flight, an automatic system was installed on the C-5A during its development phase that sensed the onset of turbulence (using, as sensors, accelerometers or wing bending strain gauges) and called for both ailerons, i.e., those on port and starboard wings, to retrim slightly upward when gust loads exceeded some level. This reduced the lift on the outboard wing sections. To preserve total lift, of course, the aircraft pitch attitude would be increased, but the total

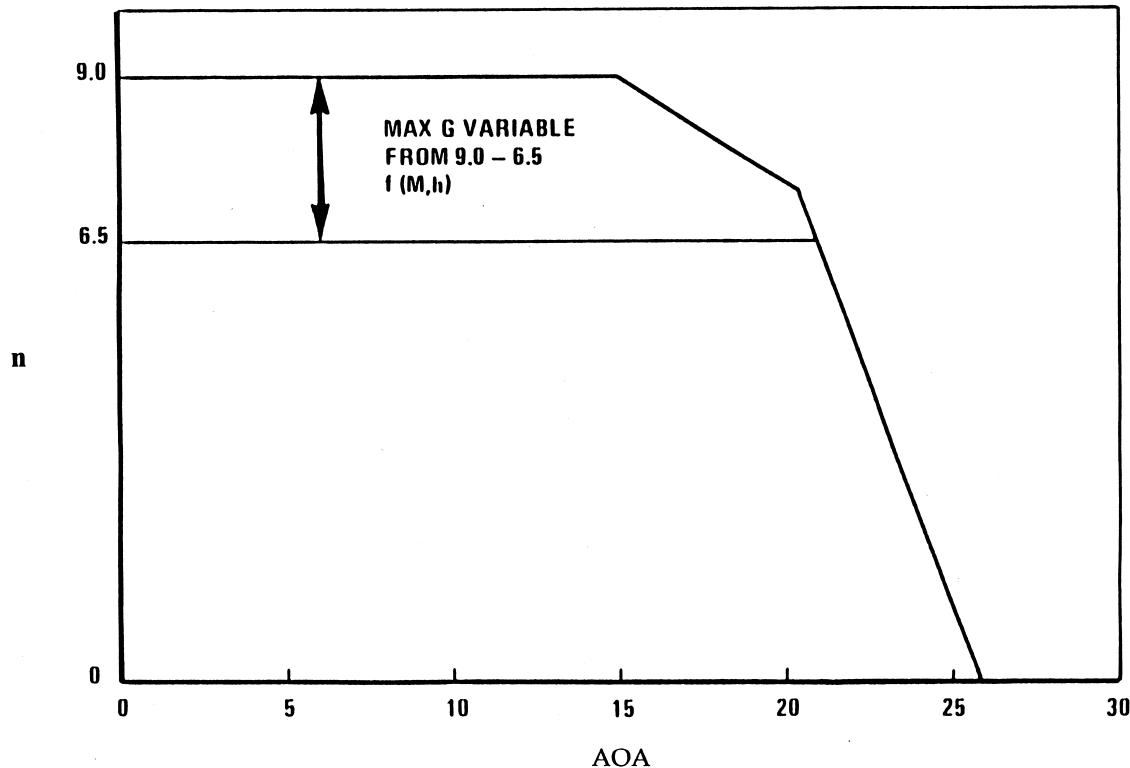


FIGURE 10 Original YF-16 AOAG Limiter Concept. (From Loewy, R. G. (2000). "Avionics: A 'New' senior partner in aeronautics." *AIAA J.* **37**(11), 1337–1354.)

effect would be a spanwise lift distribution deviating from the approximately elliptical by having less load outboard, more inboard. This load distribution reduces steady wing root bending moments and associated steady stresses because of bending, relative to that produced by elliptical distributions.

With this avionics system operative, the aircraft would have optimum cruise efficiency in smooth air, i.e., turbulence below some predetermined level, and lower aerodynamic efficiency when cruising in turbulent air. In turbulence, however, its wing root structural material would enjoy a more favorable position on the Goodman Diagram, so as to improve the wing structure's fatigue life. If such a system were found to be advantageous for an operational aircraft in the part of preliminary design concerned with wing structural weight, its designers could rely on lower combinations of steady and alternating stresses than would have existed if the more nearly elliptic spanwise lift distribution had been carried under all atmospheric conditions. Note that no quick-acting automatic actuation would be required for this spanwise lift distribution modification in rough air. However, because flight path control effectors (ailerons) are involved, only limited authority would be given the actuators. Thus, if the system failed so as not to operate, the pilot could compensate by reducing airspeed,

hence loads caused by turbulence, to levels of acceptable alternating stresses. If it failed with the ailerons in the up position, the pilot could take some range-preserving actions which might, for example, adversely affect flight time, but preserve flight safety. In either case, however, appropriate pilot action would depend on having displays that reveal the existence and nature of the failure in this system. Redundancy in this kind of automatic control system and a requirement for full or partial authority operation after a single failure, however, would not be necessary characteristics.

Load limiting as described in the F-16 example clearly requires VMS response as quick or quicker than a pilot can command. Thus, as a consequence of the level of reliability needed, all possible failures in systems components must be accounted for in design stages and adverse consequences minimized. This requires that multiple sensors and processors be provided for redundancy. It is not unusual for large commercial jet transports to have five full authority digital processors to control pitch, roll, and yaw; and to have each such computer divided into two physically separated channels. The first one, the control channel, is permanently monitored by the second one, the monitor channel. In the case of disagreement between control and monitor, the computer affected by the failure is

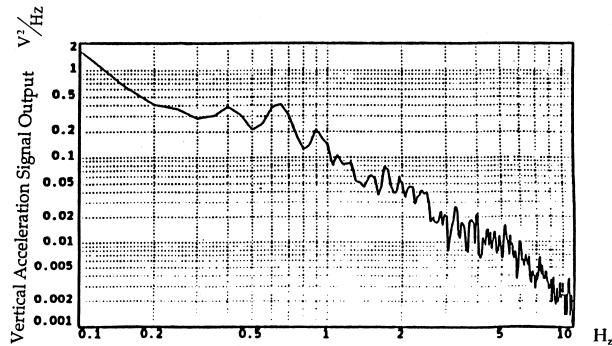


FIGURE 11 Gust frequency spectrum. (From Loewy, R. G. (2000). "Avionics: A 'New' senior partner in aeronautics." *AIAA J.* 37(11), 1337–1354.)

turned off automatically, while the computer with the next highest priority takes control. In addition, to prevent common mode failures, designers of such systems will often accept the cost penalties associated with dissimilarity to provide two types of computers.

C. Crew/Passenger Comfort Improvement

In addition to the emphasis on safety associated with the responsibility of carrying passengers, competition among airlines to win the business of these passengers motivates commercial transport airplane designers to make flight as comfortable as possible for their customers despite atmospheric turbulence. As seen in Fig. 11 (from Loewy (2000)) there are significant frequency components in atmospheric turbulence which place systems designed to ameliorate those effects above 2 Hz; and this frequency is beyond a human pilot's control input capabilities.

It is possible to take advantage of avionics systems' high frequency capabilities, to carry out AFCS, load lim-

iting and gust alleviation for comfort purposes using many of the same system components. Figure 12 (from Loewy (2000)) is a control diagram for processors in a system to provide the last of these three functions. In that figure the words "Fuselage Acceleration" indicate the pertinent sensor, and the words "Control Surface" indicate the aerodynamic effector (ailers or elevators, or both). As an indication of the kind of improvement possible using such systems, Fig. 13 (from Loewy (2000)) shows the magnitude of Gust Load Alleviation (GLA) achieved on tests. Here wing bending moment is a direct indicator of dynamic lift variations caused by continuous turbulence, which will result in passenger/crew vertical accelerations as measured by wing accelerometers and bending strain gauges and using the wing's ailerons as corrective control effectors.

D. Suppressing Servo-Aeroelastic Instabilities

Structural problems of a dynamic and/or aeroelastic nature are usually thought of as falling into either self-excited or forced vibratory classes. In aircraft, forced structural dynamic response of airframe components has many sources of excitation. Some of these are (1) gusts, i.e., atmospheric turbulence, as mentioned in Sections IV.B and C; (2) aircraft wake induced turbulence (examples include tail buffet as caused by both boundary-layer separation and by trailing vortices from rotor/propellers where such exist; (3) engine and rotor/propeller vibratory hub forces and moments; (4) rotor/propeller blade tip passage in close proximity to a fuselage; (5) transmission of gear box vibrations at tooth contact frequencies; and (6) rapid fire weapon recoil and/or muzzle pressures in military aircraft. The phenomenon known as "flutter" of wings and tail surfaces, the latter usually coupled with aft fuselage motion, and both sometimes coupled with control surface

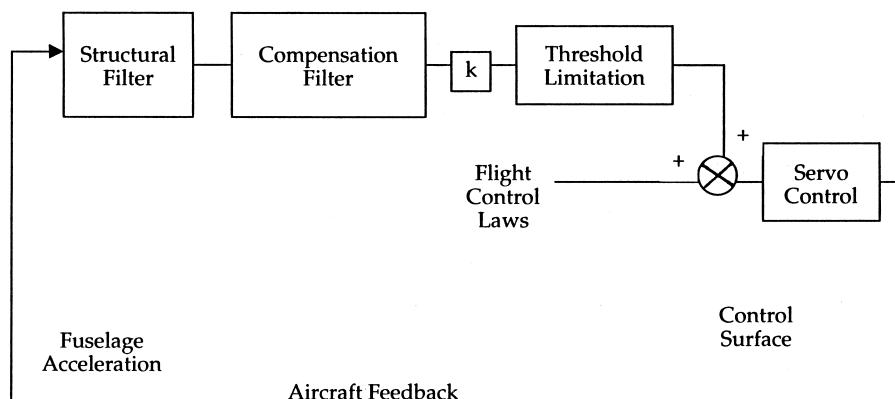


FIGURE 12 Turbulence damping: Principle of one lane (from Loewy, R. G. (2000). "Avionics: A 'New' senior partner in aeronautics." *AIAA J.* 37(11), 1337–1354.)

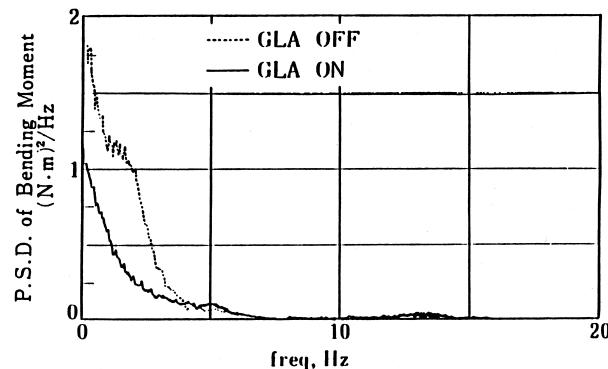


FIGURE 13 Power spectral densities of bending moment due to continuous turbulence with GLA System On & Off. (From Loewy, R. G. (2000). "Avionics: A 'New' senior partner in aeronautics." *AIAA J.* **37**(11), 1337–1354.)

deflections, is in the self-excited, i.e., aeroelastic stability class. So are the whirl-flutter and coupled rotor-wing phenomena which can be thought of as including “mechanical instabilities,” such as helicopter so-called “ground resonance” (a misnomer), as a subcase. Pilot-Induced or Pilot-Assisted Oscillations (PIO or PAO) can also be a special category of these phenomena.

The design of systems to suppress aeroelastic or aeromechanical instabilities must consider such systems as SAS, active, if such exist, because of their possible effect on the unaugmented aircraft’s structural dynamic and/or aeroelastic behavior. When a stable aeroelastic mode is destabilized by a flight path control system such as SAS, this is often called “spillover.” Aeroelastic stability (and flight path stability, for that matter) must, therefore, be assured with all systems functioning in all possible modes of operation, including partial failures. Active modification of structural and/or aeroelastic phenomena by means of avionics systems may, in any event, be thought of as acting by virtue of either reducing the (usually aerodynamic) forcing functions, generating directly opposing forces, or by introducing stiffness changes and/or additional damping into the motions crucial to the instability.

The U.S. Air Force investigated the use of avionics to reduce airframe structural design criteria to ensure aeroelastic, i.e., flutter, stability more than 20 years ago. In that research, a modified B-52 jet bomber aircraft was flown 18 km/h faster than its flutter speed. The Flutter Mode Control (FMC) system employed in that program had vertical accelerometers in pairs at four locations on the wing, which produced signals which, processed by shaping filters, drove outboard ailerons in one independent loop, sensors to surfaces, and outboard flaperons, in a second. The system was predicted to increase flutter placard speed by more than 30% by increasing damping

in and improving coupling between the structural modes active in the aeroelastic instability.

V. IMPACT OF “SMART MATERIALS”

Smart materials are thought of as those which produce electric voltages when strained (the “direct” piezoelectric effect, for example), and/or become strained when subjected to electric or magnetic fields or temperature changes. Examples of the latter behavior include piezoelectric materials (the “converse” piezoelectric effect), magnetostrictive materials, and “shape memory” alloys (which change states, and thereby dimensions or shapes, depending on temperature), respectively. The strain sensing characteristic gives piezoelectrics potential for use as sensors, the strain-inducing behavior of the last three, potential as actuators. Among the unique properties of such materials for use in avionics systems are included the possibility of distributing them throughout a structure, rather than concentrated at specific points as are such sensors as accelerometers or strain gages, or as actuators are, which drive the rotation of a control surface, for example. In some applications, very high frequency response is available, as well. Where distributed sensing has advantages, smart materials are considered for embedding as “just another fiber” in advanced filamentary composite materials, and embedding may be considered for distributed actuation too, but usually at a surface of the structure.

Using smart materials and structures techniques also holds promise for more spatially continuous variations of shapes and other properties, with the elimination of discontinuities in slope angles at the surface—often important from aerodynamic considerations.

Piezoelectric materials are particularly promising for avionics functions requiring high frequency sensing and/or actuation. For example, techniques are being developed which could be integrated into turbine engine controls, to allow turbo-machinery performance increases and reductions of chemical and/or particulate elements in their exhaust harmful to the environment. Turbine engine performance generally improves the closer their operations are taken to compressor stall. Yet stall in service operation is unacceptable. Feedback control has been shown to be capable of extending the effective stable flow range of axial compressors and rejecting persistent disturbances which, otherwise, would cause the system to incur rotating stall.

As regards achieving more efficient and “cleaner” combustion, a key consideration is in avoiding combustion instabilities. Such instabilities can involve large amplitude acoustic oscillations sufficient to cause mechanical or thermal damage, and passive approaches to avoiding

them have generally not been satisfactory. On the other hand, tests using active control systems have been shown to be promising. A pressure sensor at an upstream location in the combustor where all axial acoustic modes are expected to be significant, can provide a signal processed so as to modulate the flow of a secondary gaseous fuel stream into the combustor, with gain and phase changing with stability characteristics in real time. The secondary fuel injector actuator requires a modulation rate variable from 0 to 1500 Hz and the processor must rely on an “observer” to “identify the amplitudes, frequencies and phases of several combustor modes in real time” (Loewy, 2000). At this writing, it appears that a fuel injector suitable for such an automatic combustion instability controller, using smart materials, could be integrated within existing engine fuel-feed systems. Both the above applications of avionics components and techniques, and presumably others, to the design and operation of “intelligent turbine engines” seem highly likely, at some point in the future.

Although no applications of “smart materials” in avionics systems of any kind have appeared in aircraft presently in service use, they continue to have considerable promise: for example, to reduce structural vibration resulting from tail buffet; to increase the aircraft speed at which panels exposed to the airstream will flutter; to change the twist—and thereby improve the performance of rotor blades; and to produce lifting surfaces (wings and tails) particularly those constructed with “tailored” filamentary composite materials, which will have superior aeroelastic stability characteristics.

VI. SUMMARY

It is useful, particularly in view of the integration of avionics components into more than one functional system, to think of these systems as being of the kind that (1) change the behavior of the uncommanded airframe, (2) change the behavior of the uncommanded engines, (3) modify the pilot’s command signals, (4) provide mission-related command signals; and (5) provide mission-enabling information to the pilot. The last of these is a function that allows the pilot to fly where he/she wants to go and to use the routes he/she chooses to get there—in the case of general aviation, commercial and military transport aircraft. In the case of combat aircraft, equipment providing this function can be considered part of the “payload,” because

finding the enemy and/or neutralizing the enemy may be the fundamental reason the mission is undertaken in the first place. These broad functional categories can be listed as follows:

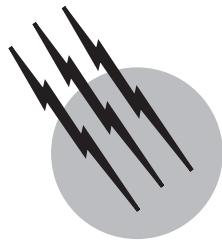
1. Changing behavior of uncommanded airframes
 - a. Flight path stability augmentations (SAS)
 - b. Aeroelastic instability suppression (FMC)
 - c. Load reduction
2. Changing behavior of uncommanded engines
 - a. Compressor stall avoidance
 - b. Emission amelioration
 - c. Noise reduction
 - d. Improving combustion efficiency
3. Acting as pilot’s associate
 - a. Threat avoidance maneuvers
 - b. Target tracking/weapon pointing
 - c. Formation flying/station keeping
 - d. Automated landings
4. Providing mission enabling pilot information
 - a. Communication
 - b. Navigation
 - c. Collision/terrain avoidance
 - d. Target/threat location

SEE ALSO THE FOLLOWING ARTICLES

AIRCRAFT INSTRUMENTS • AIRCRAFT PERFORMANCE AND DESIGN • AIRCRAFT SPEED AND ALTITUDE • AIRPLANES, LIGHT • FLIGHT (AERODYNAMICS)

BIBLIOGRAPHY

- Droste, C. S., and Walker, J. E., (1990). “A Case Study on the F-16 Fly-by-Wire Flight Control System,” AIAA, Inc., Reston, VA.
- Kayton, M., and Fried, W. R. (1997). “Avionics Navigation Systems,” 2nd ed., Wiley, New York.
- Loewy, R. G. (2000). “Avionics: A “New” senior partner in aeronautics,” *AIAA J.* **37**(11), 1337–1354.
- McClintock, F. A., and Argon, A. S., (1966). “Mechanical Behavior of Materials,” Addison-Wesley, Reading, MA.
- “Reference Data for Engineers: Radio, Electronics, Computer and Communications” (1985). 7th ed., Sams, Howard W. & Co., Indianapolis, IN.
- “Sagnac Effect,” *Review of Modern Physics* (April 1967). Vol. 39, No. 2.
- Skolnik, M. I. (1962). “Introduction to Radar Systems,” McGraw Hill, New York.



Aircraft Aerodynamic Boundary Layers

Jean Cousteix

ONERA and SUPAERO

- I. Introduction
- II. Body in Motion in a Fluid
- III. Stresses and Heat Fluxes in a Fluid
- IV. Laminar and Turbulent Flows
- V. High Reynolds Number Flows
- VI. An Example of Boundary Layers:
Falkner-Skan Solutions
- VII. Laminar-Turbulent Transition
- VIII. Turbulent Boundary Layers
- IX. Drag Reduction
- X. Concluding Remarks

GLOSSARY

Boundary layer Thin layer of viscous, possibly turbulent, flow near a wall where the velocity exhibits very fast variations normal to the wall. Boundary layers develop at high Reynolds numbers.

Diffusion of momentum (or heat) Transport of momentum (or heat) by viscosity (or thermal conductivity).

Dissipation Transformation of kinetic energy into heat due to the deformation work of viscous stresses.

Drag Component of aerodynamic forces aligned with the relative velocity between the body and the free stream.

Inviscid flow Approximation in which viscous effects are negligible.

Navier-Stokes equations Equations governing fluid flow.

Reynolds number Dimensionless number which gives the magnitude of the ratio of inertial forces to viscous forces in a flow.

Skin friction or wall shear stress Stress at the wall due to viscosity. The corresponding force applied to the body is parallel to the wall.

Stress tensor Second-order tensor defined at any point in a flow and determining a force applied to a surface element bounding a volume of fluid. The dimension of stress tensor components is a force per unit surface.

Viscosity Property of a fluid by which momentum diffuses in a flow. Viscosity smooths inhomogeneities of momentum.

PHYSICAL and theoretical aspects of boundary layer flow are presented, stressing important effects such as friction drag and separation. Laminar-turbulent transition and turbulence effects are also described. These phenomena have a strong influence on the development of boundary layers but their understanding and modeling are still largely open questions. Finally, drag-reduction techniques by means of riblets and laminarization are presented.

I. INTRODUCTION

The aerodynamic flow around the wing of a civil aircraft in cruise conditions is characterized by a very large Reynolds number on the order of several tens of millions. The Reynolds number can be interpreted as the order of magnitude of the ratio of inertial forces to viscous forces. At high Reynolds numbers, viscous effects are negligible. This assumption is valid almost everywhere in the flow except close to the walls where viscosity is effective. Viscous effects are also important in wakes and jets.

With appropriate hypotheses, the Navier-Stokes equations are drastically simplified and the boundary layer theory makes flow analysis much easier. Viscosity is responsible for friction drag or separation, which are among the most important characteristics of a wing. These remarks explain the key role played by boundary layers in aerodynamics.

As a general rule, the boundary layer flow is not laminar all along its development on a wing. *Laminar-turbulent transition* and *turbulence* take place more or less early according to the conditions of boundary layer development. These phenomena have a strong influence on the behavior of boundary layers. At the present time, the understanding of these phenomena is not complete and no definitive model exists.

Basic concepts of aerodynamics and fluid mechanics are reviewed in the first sections of this paper. The rest of the presentation is devoted to a description of the physical and theoretical aspects of the following topics: boundary layer concept, laminar-turbulent transition, turbulence effects, and drag reduction techniques.

II. BODY IN MOTION IN A FLUID

In standard aerodynamics, air is a fluid considered as a continuum. The concept of *fluid particle* is defined as a volume of fluid large enough to contain a great number of gas molecules and small enough to be characterized by averaged quantities which are uniform over the fluid particle. A *fluid flow* is characterized by thermodynamic properties such as pressure, temperature, density,

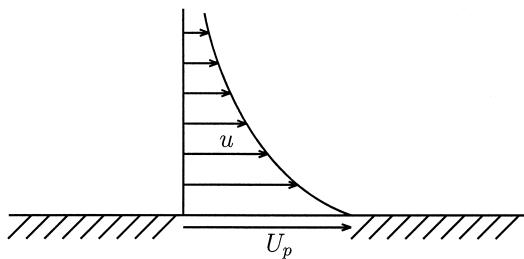


FIGURE 1 Fluid entrainment by a flat plate moving parallel to itself.

etc. A fluid particle is entrained at the local velocity of the flow.

In an *inviscid flow*, excluding the possibility of modifications of the molecules' structure at high temperature due to chemical reactions, for example, the spatial or temporal variations of the thermodynamic properties of the fluid are associated with velocity variations or with inhomogeneities of the initial or boundary conditions.

In a *real flow*, other phenomena modify the flow properties. Fluids, like solids, are *heat conductors*. *Thermal conductivity* is a property of the fluid that tends to even out the temperature differences in the fluid. If a volume of hot fluid is in contact with a volume of colder fluid, thermal conductivity creates a heat flux which transfers heat from the hot fluid to the cold fluid. Fluids have another property—*viscosity*—which tends to smooth the momentum differences in the fluid, which implies a smoothing of velocity differences. For example, a body with a flat surface moving parallel to its surface entrains the fluid in its neighborhood (Fig. 1). Momentum diffuses in the fluid. Momentum diffusion is similar to the heat diffusion which would take place if fluid at rest were in contact with a hotter (or colder) plate. As shown in Fig. 1, the fluid velocity is equal to the velocity of the body along the fluid-body interface. In standard aerodynamics, the temperature of the fluid and of the solid are also identical along the fluid-solid interface.

A. Forces in the Fluid

Let us consider a fluid volume V surrounded by fluid and limited by a surface Σ (Fig. 2). Let $d\sigma$ be an element of this surface. First, let us assume that the flow is inviscid. Along $d\sigma$ the volume V is submitted to a force normal to $d\sigma$ which is called the *pressure force*. The pressure force points towards the inside of volume V . In a viscous flow, the pressure force is also present but another force is exerted on volume V . This additional force due to viscosity is not generally normal to $d\sigma$. In a viscous flow, the fluid in V can also exchange heat with the outer fluid through $d\sigma$.

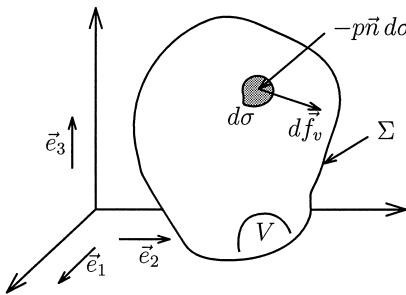


FIGURE 2 Local forces on the surface of a volume of fluid. The vector \vec{n} is a unit vector normal to $d\sigma$ and pointing outwards from V .

B. Forces on the Body

Now, let us consider the surface of a solid in contact with a fluid (Fig. 3). In an inviscid flow, along a surface element $d\sigma$ the fluid exerts a pressure force on the solid; this pressure force is normal to $d\sigma$. In a viscous flow, an additional force due to viscosity is tangential to $d\sigma$; heat transfer can also occur between the fluid and the solid through $d\sigma$.

In aerodynamics, it is often convenient to consider that the fluid is moving around the body at rest. Indeed, it is shown that if the motion is stationary (independent of time), the flow around the body moving through the fluid is equivalent to the flow around the body at rest immersed in a moving fluid. Only the *relative* motion of the body with respect to the fluid is significant. In particular, in a wind tunnel, the body is at rest and the fluid moves around it.

An airfoil is a two-dimensional body designed to produce a lift. The relative velocity of the fluid with respect to the airfoil measured far ahead of the airfoil is the free-stream velocity V_∞ (Fig. 4). The chord c of the airfoil is the distance between the leading edge and the trailing edge. The chord line connecting the leading edge and the trailing edge forms an angle α —the angle of attack or the incidence—with the free-stream velocity. The total force exerted on the airfoil is obtained by integrating the pressure forces and the viscous forces over the body's surface. The aerodynamic forces are resolved into the lift L and the drag D acting perpendicular and parallel to V_∞ , respectively.

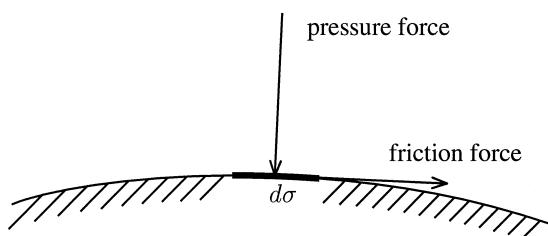


FIGURE 3 Forces applied to a surface element of a body in relative motion in a fluid.

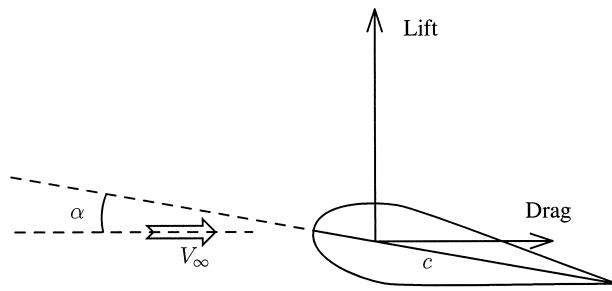


FIGURE 4 Aerodynamic forces on an airfoil.

III. STRESSES AND HEAT FLUXES IN A FLUID

A. Modeling of Stresses and Heat Fluxes

Frequently, the flow is described by the Euler variables, i.e., by the velocity field at any point in space and at any time. The description is completed by the density, pressure, and temperature fields.

The velocity gradient is a second-order tensor whose components in an orthonormal axis system are $\frac{\partial u_i}{\partial x_j}$. The velocity gradient is broken down into a symmetric tensor s_{ij} ($s_{ij} = s_{ji}$) and a skew-symmetric tensor r_{ij} ($r_{ij} = -r_{ji}$):

$$\frac{\partial u_i}{\partial x_j} = s_{ij} + r_{ij}$$

with

$$s_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \quad \text{and} \quad r_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} - \frac{\partial u_j}{\partial x_i} \right)$$

The rate of strain tensor s_{ij} defines the deformation of a small volume of fluid, whereas tensor r_{ij} defines the rotation of this volume. Indices i and j can take the values 1, 2, or 3; x_1, x_2, x_3 are x, y, z , respectively; and the corresponding velocity components are u, v, w .

Let V be a control volume of fluid (Fig. 2). The force exerted on the control volume over its surface element $d\sigma$ is

$$d\vec{f} = d\vec{f}_{\text{pressure}} + d\vec{f}_{\text{viscosity}}$$

where the pressure force and the viscous force are given by:

$$d\vec{f}_{\text{pressure}} = -p \delta_{ij} n_j d\sigma \vec{e}_i \quad (1)$$

$$d\vec{f}_{\text{viscosity}} = \tau_{ij} n_j d\sigma \vec{e}_i \quad (2)$$

The Einstein notation is used in the above formula. The repetition of an index in a term implies a summation. For example, $\tau_{ij} n_j$ is equivalent to $\sum_{j=1,3} \tau_{ij} n_j$ (i.e., $\tau_{i1} n_1 + \tau_{i2} n_2 + \tau_{i3} n_3$).

The quantities p and τ_{ij} are called *stresses*. They have the dimension of a force per unit area. The stresses p and τ_{ij} are defined at any point in space, whereas $d\vec{f}_{\text{pressure}}$ and $d\vec{f}_{\text{viscosity}}$ depend on the surface element via the unit normal \vec{n} .

In relations 1 and 2, p is the pressure, τ_{ij} is the viscous stress tensor, δ_{ij} is the second-order unit tensor ($\delta_{ij} = 1$ if $i = j$, $\delta_{ij} = 0$ otherwise), n_j are the components of the unit vector orthogonal to $d\sigma$ and pointing outwards from volume V , and \vec{e}_i are the unit vectors of the reference orthonormal axis system. Formulas (1) and (2) express that the pressure force is normal to $d\sigma$, whereas the viscous force can have a tangential component. It is shown that τ_{ij} is a symmetric tensor (i.e., $\tau_{ij} = \tau_{ji}$).

The amount of heat received by the control volume V through $d\sigma$ per unit time is

$$dQ = -\phi_j n_j d\sigma$$

where ϕ_j are the components of the heat flux vector.

As a good approximation, air is considered as a Newtonian fluid, i.e., the viscous stress tensor is expressed by a *linear* function of the rate of strain tensor:

$$\tau_{ij} = 2\mu s_{ij} + \kappa \frac{\partial u_l}{\partial x_l} \delta_{ij}$$

where μ and κ are viscosity coefficients. This relationship assumes that r_{ij} does not create any viscous stress.

For most practical purposes in standard aerodynamics, Stokes' hypothesis, $2\mu + 3\kappa = 0$, is valid. Then, the viscous stress tensor is given as:

$$\tau_{ij} = 2\mu s_{ij} - \frac{2}{3}\mu \frac{\partial u_l}{\partial x_l} \delta_{ij}$$

Stokes' hypothesis implies that an isotropic compression or dilatation does not create any viscous stress.

The dynamic viscosity coefficient μ can be obtained from the kinetic theory of gases or experimentally. For air, a good representation of μ is obtained from Sutherland's formula:

$$\mu = \mu_0 \sqrt{\frac{T}{T_0}} \frac{1 + S/T_0}{1 + S/T}$$

Where $\mu_0 = 1.711 \cdot 10^{-5}$ Pl (the unit is 1 Poiseuille = $1 \text{ kg m}^{-1} \text{ s}^{-1}$), $T_0 = 273 \text{ K}$, $S = 110.4 \text{ K}$. Sometimes, the kinematic viscosity coefficient ν is used, $\nu = \frac{\mu}{\rho}$.

The Fourier law is used to express the heat flux vector:

$$\phi_j = -\lambda \frac{\partial T}{\partial x_j}$$

where the thermal conductivity λ is related to the viscosity through the Prandtl number P :

$$P = \frac{\mu c_p}{\lambda}$$

For air, the Prandtl number is considered as a constant $P = 0.725$ for temperatures below 1500 K, and the specific heat at constant pressure c_p is also a constant, $c_p = 1005 \text{ J kg}^{-1} \text{ K}^{-1}$.

Viscosity and thermal conductivity are *thermodynamic properties* of the fluid. Even when the fluid is at rest, these properties are defined. The behavior of the flow is influenced by the *effects of viscosity and thermal conductivity*—nonzero viscous stresses and heat fluxes—which exist only when there is a nonzero rate of strain and temperature gradient.

B. Navier-Stokes Equations

The motion of the flow is governed by the Navier-Stokes equations. These equations comprise

- The continuity equation which expresses the conservation of mass.
- The momentum equation which expresses Newton's second law, $\vec{F} = m\vec{\gamma}$: the acceleration of a fluid particle is due to external forces applied to it.
- The energy equation which expresses the first principle of thermodynamics: the variation of the total energy (internal energy plus kinetic energy) of a fluid particle is due to the heat exchanged with the surrounding medium and the work done by the external forces.

These equations are completed by a state equation. For a perfect gas, the state equation is

$$\frac{p}{\rho} = \frac{\mathcal{R}}{\mathcal{M}} T$$

where \mathcal{R} is the universal constant of perfect gases, $\mathcal{R} = 8.3145 \text{ J mol}^{-1} \text{ K}^{-1}$; and \mathcal{M} is the molar mass of the gas. For air, we have $R = \frac{\mathcal{R}}{\mathcal{M}} = 287.1 \text{ J kg}^{-1} \text{ K}^{-1}$.

The Navier-Stokes equations are

$$\text{Continuity equation } \frac{\partial \rho}{\partial t} + \text{div}(\rho \mathbf{u}) = 0 \quad (3)$$

Momentum equation

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{grad} \mathbf{u}) \cdot \rho \mathbf{u} = \text{div}[-p \underline{\underline{\delta}} + \underline{\underline{\tau}}] \quad (4)$$

Energy equation

$$\rho \frac{\partial h_i}{\partial t} + \rho \mathbf{u} \cdot \mathbf{grad} h_i = \frac{\partial p}{\partial t} + \text{div}(\underline{\underline{\tau}} \cdot \mathbf{u} - \phi) \quad (5)$$

where \mathbf{u} is the velocity vector, $\underline{\underline{\tau}}$ is the viscous stress tensor, $\underline{\underline{\delta}}$ is the second order unit tensor ϕ is the heat flux vector, div is the divergence operator, \mathbf{grad} is the gradient operator, and \cdot denotes the dot product. In the above equations, h_i is the *stagnation enthalpy*:

$$h_i = h + \frac{|\mathbf{u}|^2}{2}$$

where h is the *static enthalpy*. For an ideal gas (the specific heats at constant pressure c_p and at constant volume c_v are constants), the static enthalpy is related to the temperature by $h = c_p T$.

A typical term representing the inertial forces is $\rho u \frac{\partial u}{\partial x}$. Let L be a length scale of the flow: this means that u has a significant variation of order V_0 over a length L . It is assumed that the order of u is V_0 and that the order of ρ is ρ_0 . Then the order of $\rho u \frac{\partial u}{\partial x}$ is $\frac{\rho_0 V_0^2}{L}$. A typical term representing the viscous forces is $\frac{\mu}{\delta x} (\mu \frac{\partial u}{\partial x})$. The order of magnitude of this term is $\frac{\mu_0 V_0}{L^2}$. Then the order of the ratio of the inertial forces to the viscous forces is the *Reynolds number*:

$$R_L = \frac{\rho_0 V_0 L}{\mu_0}$$

On a transport aircraft, the Reynolds number is very large compared to unity. The Reynolds number based on the free-stream conditions and on the average chord length is several tens of millions. Indeed, the very large Reynolds number is a very important characteristic of aerodynamic flows. From the above interpretation, this would mean that the viscous forces are negligible compared with the inertial forces. This means it would not be necessary to take the viscous forces into account and the inviscid flow approximation could be applied. This hypothesis is valid almost everywhere except in the vicinity of the walls where the velocity vanishes and where the *local* Reynolds number can be small. Near walls, the velocity exhibits very fast variations normal to the surface. The strongly sheared flow generates viscous forces which are of the same order as the inertial forces. The near wall flow structure is called the *boundary layer* (Fig. 5). Viscous forces also play an important role in wakes which develop downstream of a wing when the upper surface and lower surface boundary

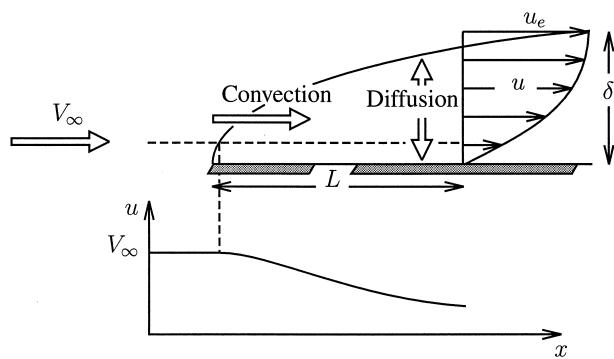


FIGURE 5 Typical velocity variations in a boundary layer. The shear flow generates viscous forces parallel to the wall and an associated diffusion of momentum normal to the wall.

layers merge or in a jet which is formed downstream of an engine.

IV. LAMINAR AND TURBULENT FLOWS

A. Properties of Turbulence

Most often, on an aircraft, the boundary layers, wakes, and jets are not laminar but *turbulent*. No precise definition of a turbulent flow exists, but some important characteristic properties are

- The characteristic Reynolds number of turbulent flows is large compared to unity.
- Turbulence is a very efficient mixer. Momentum transfer by viscosity or heat transfer by conductivity are slow processes, whereas turbulent motions are much more efficient in exchanging momentum and heat between parts of the flow that contain different amounts of momentum and heat. On the average, a turbulent flow is more uniform than a laminar flow.
- The velocity field is three dimensional and rotational. Statistically, the flow can be two dimensional or even one dimensional, but the velocity fluctuations are three dimensional. This property has an important consequence on the rotational characteristic of the flow because the dynamics of vorticity are very different if the velocity field is two or three dimensional. In a two-dimensional incompressible flow, vorticity develops under the influence of convection and diffusion due to viscosity. In a three-dimensional flow, vorticity can be created or destroyed by the action of the rate of strain tensor s_{ij} . In particular, stretching of the vorticity line increases vorticity; this process plays an important role in turbulence.
- Turbulence is a dissipative phenomenon. The deformation work done by the viscous stresses transforms kinetic energy into heat. This process also exists in laminar flows but is more efficient in turbulent flows.
- Turbulence is described by Navier-Stokes equations. A turbulent field comprises a large range of length scales. For example, if turbulence is created in a cubic room, the largest scales of turbulence can be the dimension of the room. If the characteristic Reynolds number of these large-scale motions is large compared to unity, these structures behave as if the flow were inviscid. Due to the stretching process, smaller scale turbulent motions exist and energy is transferred to them. This process is repeated until the characteristic Reynolds number of these motions is near unity. The

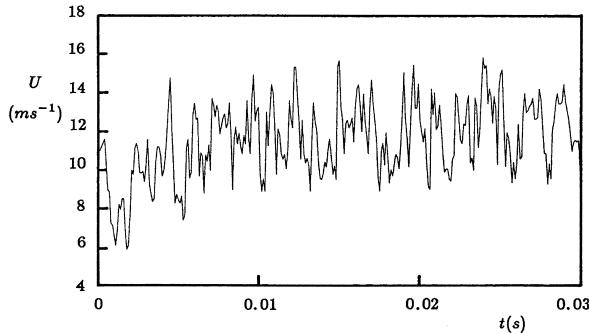


FIGURE 6 Time variation of the velocity in a point of a turbulent boundary layer. The signal looks random.

viscosity is efficient in transforming all the kinetic energy into heat and no smaller motions exist. These dissipative scales are called the *Kolmogorov scales*. For realistic cases, these scales are very different from the molecular scales and the turbulent field can be considered as a continuum.

- Turbulence is a nonlinear phenomenon due to the nonlinearity of the Navier-Stokes equations. This property makes the problem very difficult.
- Turbulent flows are not predictable. Even if it were possible to solve the Navier-Stokes equations perfectly (using numerical methods), the least error would contaminate the flow after a finite time. In a given point of space, it is impossible to know the variation of the velocity after this time.
- Turbulence has a random character. At first sight, the variation of the velocity as a function of time in a given point of space looks random (Fig. 6); however, turbulence consists of coherent structures (Fig. 7). The correlation of the velocities in two points of space is nonzero when the separation distance is not too large. The coherent structures have finite dimensions in space and time.

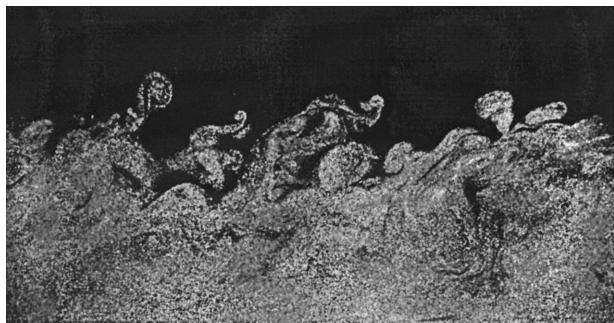


FIGURE 7 Visualization showing coherent structures in a turbulent boundary layer flow. [Courtesy of M. Stanislas.]

B. Averaged Navier-Stokes Equations

The fluctuations of the flow characteristics are described by the Navier-Stokes equations, but their range of length and time scales is very large. The storage space and CPU time required to resolve this spectrum of fluctuations in practical situations are not available and will not be available in the foreseeable future. For research purposes, in simple flow configurations, *direct numerical simulations* consisting of solving the three-dimensional, unsteady Navier-Stokes equations are a very valuable tool. Other techniques consist of calculating the largest structures in the flow and modeling their interaction with the smaller structures. *Large eddy simulations* are a field of intense research.

For practical applications, the *averaged Navier-Stokes equations* are used. The flow characteristics are broken down into an average component and a fluctuating component. In an incompressible flow, the average is defined as a statistical or ensemble average. For example, the average longitudinal component of velocity is

$$\langle u \rangle = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N u_i$$

where u_i are samples taken in different repetitions of the same flow (the same experiment is repeated many times). Theoretically, the number of samples is infinite. In an experiment, this number is large but not infinite. In the same way, the pressure is broken down into average and fluctuating components. In a compressible flow, the best way to define an average is not a trivial matter. Very often, mass-weighted or Favre averaging is used. The average velocity \tilde{u} is

$$\tilde{u} = \frac{\langle \rho u \rangle}{\langle \rho \rangle}$$

where $\langle \bullet \rangle$ denotes a statistical average. Statistical averages are used for the density and pressure. The advantage of this breakdown is that the continuity equation has the standard form and it is possible to work on closed systems as is done usually.

To obtain the equations for the average flow, the Navier-Stokes equations are averaged, taking into account the breakdown of the flow. For the sake of simplicity, it is sufficient here to give the equations for incompressible flows:

$$\operatorname{div}(\mathbf{u}) = 0 \quad (6)$$

$$\rho \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{grad} \mathbf{u}) \cdot \rho \mathbf{u} = \operatorname{div}[-p \underline{\delta} + \underline{\tau} + \underline{R}] \quad (7)$$

In these equations, all the quantities are averaged quantities but the sign $\langle \bullet \rangle$ has been omitted when there is no confusion.

The above equations—the *Reynolds equations* or *averaged Navier-Stokes equations*—show that turbulence affects the average flow only through the tensor $\underline{\underline{R}}$ which is the *Reynolds stress tensor*. In an orthonormal axis system, the components of this tensor are

$$R_{ij} = -\rho \langle u'_i u'_j \rangle$$

The Reynolds stress tensor components come from nonzero velocity correlations. This tensor is symmetric since $\langle u'_i u'_j \rangle = \langle u'_j u'_i \rangle$. The Reynolds stresses are grouped with the viscous stresses and appear as *apparent turbulent stresses*. In fact, these terms occur when the nonlinear convection term of the momentum equation is averaged:

$$\langle u_i u_j \rangle = \langle u_i \rangle \langle u_j \rangle + \langle u'_i u'_j \rangle$$

The term $\langle u'_i u'_j \rangle$ has been put in the right-hand side of the averaged momentum equation to express the equivalence between momentum and force.

The turbulent kinetic energy per unit mass is

$$k = \frac{\langle u'_i u'_i \rangle}{2} = \frac{\langle u'^2 \rangle + \langle v'^2 \rangle + \langle w'^2 \rangle}{2}$$

The Reynolds stresses represent the influence of turbulence on the average flow. There is no obvious way of relating $\langle u'_i u'_j \rangle$ to the mean flow properties, and Eqs. (6) and (7) are not closed. The Reynolds stresses are additional unknowns which must be modeled in order to solve the Reynolds equations. The *turbulence modeling problem* is a major difficulty which has not yet been solved to describe all turbulent flows with a unique model. For certain classes of flows, turbulence models are able to give reasonable results.

V. HIGH REYNOLDS NUMBER FLOWS

A. Boundary Layer Hypotheses

Viscous forces are negligible compared with inertial forces when the Reynolds number is large. For the flow around an airfoil, this conclusion is nearly correct except that it leads to important contradictions with physics. If viscous effects were negligible everywhere in the flow, an airfoil at low speed would have no drag. Common experience tells us that this result is wrong. A second drawback would be that the continuity of velocity or temperature along the interface between a solid and a fluid would be violated.

Prandtl introduced the idea that, in a laminar flow, viscosity should affect the flow even if the Reynolds number is very large. The viscous effects are confined in a thin layer near the wall—the *boundary layer*—where the viscous forces are of the same order as the inertial forces. This hypothesis is not at variance with the meaning of the Reynolds number. Simply, the significant Reynolds num-

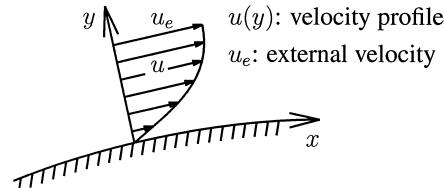


FIGURE 8 Boundary layer coordinate system.

ber for the boundary layer is based on an appropriate length scale—its thickness—which is very small compared with a characteristic length of the body. The boundary layer hypothesis also requires the boundary layer thickness to be small compared to the wall curvature radius.

According to the boundary layer hypothesis, the main viscous forces are parallel to the wall and diffusion occurs essentially in a direction normal to the wall (Fig. 5). Another important consequence of the boundary layer hypothesis is that the static pressure variations across the boundary layer are negligible.

The boundary layer theory can also be applied to a turbulent flow. Below, the boundary layer equations are written for a turbulent flow but, for the sake of simplicity, they are given for a two-dimensional steady flow (the *mean flow* is two dimensional and steady). The x -coordinate follows the contour of the body and the y -coordinate is normal to it (Fig. 8). The compressible flow equations are

$$\text{Continuity equation } \frac{\partial \rho u}{\partial x} + \frac{\partial \rho v}{\partial y} = 0 \quad (8)$$

x -Momentum equation

$$\rho u \frac{\partial u}{\partial x} + \rho v \frac{\partial u}{\partial y} = -\frac{\partial p}{\partial x} + \frac{\partial}{\partial y} \left(\mu \frac{\partial u}{\partial y} - \rho \langle u' v' \rangle \right) \quad (9)$$

$$y\text{-Momentum equation } 0 = \frac{\partial p}{\partial y} \quad (10)$$

Energy equation

$$\begin{aligned} \rho u \frac{\partial h_i}{\partial x} + \rho v \frac{\partial h_i}{\partial y} &= \frac{\partial}{\partial y} \left[u \left(\mu \frac{\partial u}{\partial y} - \rho \langle u' v' \rangle \right) \right. \\ &\quad \left. + \lambda \frac{\partial T}{\partial y} - \rho \langle h' v' \rangle \right] \end{aligned} \quad (11)$$

These equations are completed by the state equation:

$$\frac{p}{\rho} = \frac{\mathcal{R}}{\mathcal{M}} T$$

In the x -momentum Equation (9), the viscous stress $\mu \frac{\partial u}{\partial y}$ is associated with an apparent turbulent stress $-\rho \langle u' v' \rangle$. Equation (9) expresses that the acceleration represented by the left-hand side is due to the pressure forces and to viscous and turbulent stresses (Fig. 9). In the energy

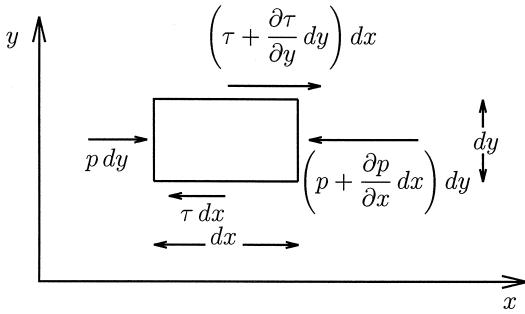


FIGURE 9 Streamwise forces in a boundary layer. The stress τ is the sum of a viscous stress $\mu \frac{\partial u}{\partial y}$ and an apparent turbulent stress $-\rho(u'v')$.

Equation (11) the heat flux $-\lambda \frac{\partial T}{\partial y}$ is associated with an apparent turbulent heat flux $\rho(v'h')$. The contribution of viscosity and thermal conductivity is to diffuse momentum and energy in the y -direction. By extension, the effect of turbulent terms is called turbulent diffusion. The above equations are written assuming that the ensemble averages are identical to mass-weighted averages.

B. Boundary Layer Characteristics

The boundary layer is associated with important characteristics such as the skin-friction coefficient and the displacement thickness. The *skin-friction coefficient* is a dimensionless quantity defined from the wall shear stress:

$$c_f = \frac{\tau_w}{\frac{1}{2} \rho_e u_e^2}$$

where ρ_e and u_e are the values of the density and of the longitudinal velocity at the edge of the boundary layer. The wall shear stress is

$$\tau_w = \left(\mu \frac{\partial u}{\partial y} \right)_{y=0}$$

The *displacement thickness* is introduced when comparing the real flow and an equivalent inviscid flow (Fig. 10). Along the boundary layer edge $y = \delta$, the flow characteristics are the same in the real flow and in the equivalent

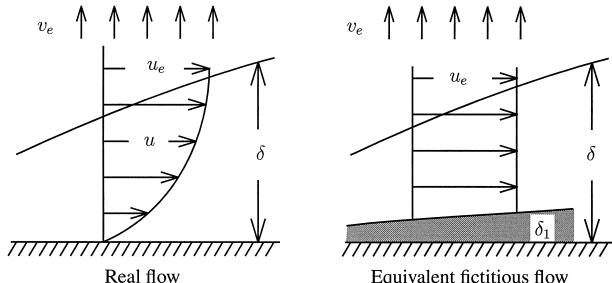


FIGURE 10 Displacement thickness.

flow. In this way, the boundary conditions for the inviscid flow in the region $y > \delta$ are the same in both flows; this ensures that the flows are identical. In the region $y < \delta$ an equivalent (fictitious) inviscid flow is defined as a continuation of the inviscid flow, and the characteristics of this fictitious flow are invariant along a normal to the wall. In order to have the same mass flow in the real flow and in the fictitious flow, the latter is limited to a region $\delta_1 < y < \delta$. The equality of mass flow gives:

$$\int_0^\delta \rho u dy = \int_{\delta_1}^\delta \rho_e u_e dy = \int_0^\delta \rho_e u_e dy - \rho_e u_e \delta_1$$

from which the definition of the displacement thickness δ_1 is

$$\delta_1 = \int_0^\delta \left(1 - \frac{\rho u}{\rho_e u_e} \right) dy$$

As the mass flow is the same in the real flow and in the fictitious flow, the displacement surface $y = \delta_1$ is a stream surface for the equivalent inviscid flow. *The displacement thickness thus represents the distance by which the body should be displaced in order to represent the boundary layer effects in the equivalent inviscid flow.*

Another way to represent the boundary layer effects in an equivalent inviscid flow is to define a blowing velocity at the wall (Fig. 11):

$$v_b = \frac{d(u_e \delta_1)}{dx}$$

Then the equivalent inviscid flow is defined from the wall to infinity, but at the wall the boundary condition is now a blowing velocity, v_b .

The momentum thickness θ and the shape factor H are also defined:

$$\theta = \int_0^\delta \frac{\rho u}{\rho_e u_e} \left(1 - \frac{u}{u_e} \right) dy; \quad H = \frac{\delta_1}{\theta}$$

The momentum thickness is the thickness which is added to the displacement thickness in order to have the same flux of momentum in the real flow and in the fictitious flow.

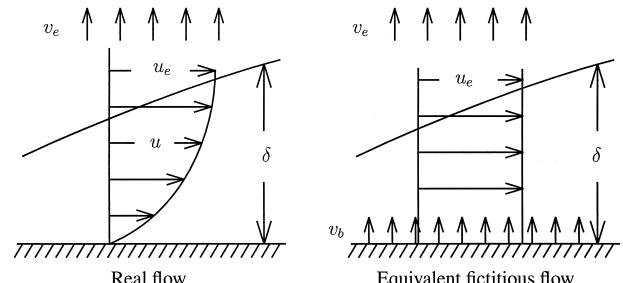


FIGURE 11 Blowing velocity.

C. Equivalence with Navier-Stokes Equations

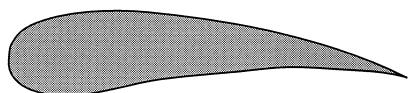
At high Reynolds numbers, the Navier-Stokes equations are equivalent to a system consisting of the inviscid flow equations and the boundary layer equations, the two systems being coupled by the displacement thickness or the blowing velocity. From the standard boundary layer theory, the flow around an airfoil can be calculated according to the following sequence (Fig. 12):

1. An inviscid flow calculation is performed around the physical airfoil. In this calculation, the surface of the airfoil is assumed to be impermeable.
2. A boundary layer calculation is performed using as an input the pressure obtained in the first step at the wall of the airfoil.
3. A new inviscid flow calculation is performed. During this step, the boundary layer effects are accounted for. If the displacement thickness concept is used, the body is displaced by a distance equal to δ_1 (Fig. 12). If the blowing velocity concept is used, the boundary condition at the wall is the value of v_b determined from the boundary layer calculation.

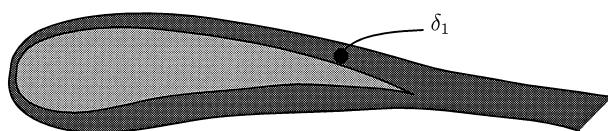
This method works as long as the boundary layer has not separated. When separation occurs, the concept of boundary layer can still be used but the sequence described above can no longer be used and the solution is more complicated.

D. Solution of Boundary Layer Equations

The boundary layer equations are solved by numerical methods. According to the standard theory, the pressure is known (from the solution of inviscid equations as mentioned above). The boundary layer equations form a system of parabolic equations. The mathematical nature of this system is similar to the heat equation. Without reversed flow, perturbations propagate in the downstream



Step 1: Calculation of the inviscid flow around the real airfoil.



Step 3: Calculation of the inviscid flow around the displaced body.

FIGURE 12 Viscous-inviscid interaction.

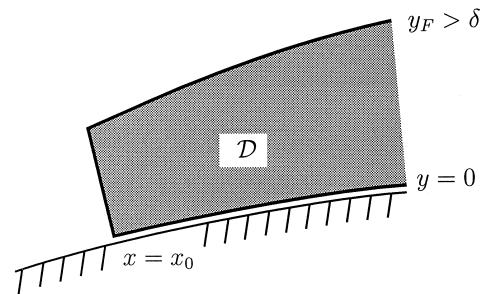


FIGURE 13 Boundary layer calculation domain.

direction along the streamlines, and they also propagate along normals to the wall. If the problem is restricted to incompressible, two-dimensional, steady flows, the boundary conditions are

- At an initial station $x = x_0$, the velocity profile is prescribed $u = u(x_0, y)$.
- At the wall, no slip conditions are prescribed: $u = 0$, $v = 0$.
- At the boundary layer edge, the velocity is equal to the inviscid velocity $u = u_e$. The external velocity u_e is related to the pressure by the Bernoulli equation:

$$\frac{dp}{dx} = -\rho_e u_e \frac{du_e}{dx}$$

According to their mathematical nature, the boundary layer equations are solved step by step in the downstream direction using marching methods from station $x = x_0$ (Fig. 13). In practice, the initial station $x = x_0$ is the stagnation point of the airfoil where local solutions are known (Section VI).

VI. AN EXAMPLE OF BOUNDARY LAYERS: FALKNER-SKAN SOLUTIONS

Particular solutions of boundary layer equations exist when the external velocity is

$$u_e = kx^m$$

where k and m are constants. Such a flow would be obtained at the wall of a wedge in an inviscid flow (Fig. 14). The opening angle of the wedge is $\beta\pi$ with:

$$\beta = \frac{2m}{m+1}$$

The variables ξ and η are introduced:

$$\xi = x; \quad \eta = \frac{y}{x} \sqrt{\frac{m+1}{2}} \frac{u_e x}{v}$$

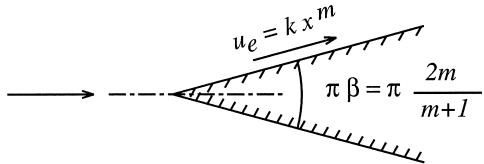


FIGURE 14 Inviscid flow on a wedge.

It can be shown that the solution to the boundary layer equations is such that:

$$\frac{u}{u_e} = f'(\eta)$$

and the momentum boundary layer equation is

$$f''' + ff'' + \beta(1 - f'^2) = 0$$

with $f = \int_0^\eta f'(\xi) d\xi$, $f'' = \frac{df'}{d\eta}$, $f''' = \frac{d^2f'}{d\eta^2}$. The boundary conditions are

$$f = 0 \text{ and } f' = 0 \text{ when } \eta = 0,$$

$$f' = 1 \text{ when } \eta \rightarrow \infty$$

The boundary layer equation is now an ordinary differential equation instead of a partial differential equation as were the original equations. The Falkner-Skan solutions are *self-similar*. The solution at a station x_2 can be deduced from the solution at any another station x_1 . The transformation rule is simply through the use of the variable η : the velocity profile $\frac{u}{u_e}$ expressed as a function of the physical variable y at a given station is obtained from the velocity profile at another station by an affinity transformation on the distance to the wall. This property transforms the partial differential equations into an ordinary differential equation for f' .

The numerical solution of the Falkner-Skan equation is shown in Fig. 15. Each velocity profile corresponds to a given value of β (i.e., a given value of the wedge angle). When $\beta > 0$, the flow is accelerated (i.e., the external velocity increases in the downstream direction); when $\beta < 0$, the flow is decelerated. The case $\beta = 0$, studied by Blasius, corresponds to the flow over a flat plate. According to the boundary layer jargon, a flat plate is an (infinitely) thin body whose surface is parallel to the free-stream velocity. Then, the pressure gradient in the flow on a flat plate is zero.

When $m = 1$, $u_e = kx$, the flow corresponds locally to the flow in the vicinity of the stagnation point on an airfoil. Then, the solution for $m = 1$ can be taken as an initial condition at the stagnation point for boundary layer calculations.

The pressure gradient has a great influence on the shape of the velocity profile and on the thickening of the boundary layer. The results show that thickening is faster when

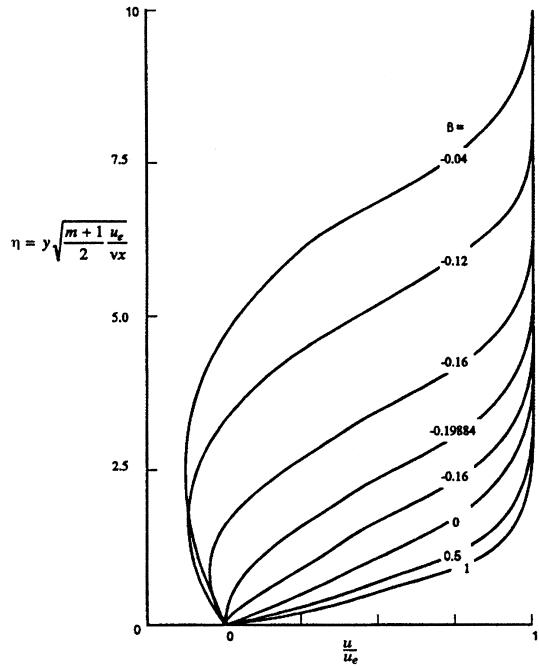


FIGURE 15 Falkner-Skan self-similar solutions.

the pressure gradient is positive. In addition, the corresponding velocity profile has an inflection point. This is a condition of instability of the velocity profile when a small perturbation is introduced in the boundary layer. Experimental results show that, in a two-dimensional flow, laminar-turbulent transition occurs earlier when the pressure gradient is positive.

When $\beta \simeq -0.19884$, the slope of the velocity profile at the wall is zero (i.e., the skin friction is zero). In a two-dimensional flow, this is characteristic of the onset of *separation*. Let Δx be the extent of a boundary layer element limited by two planes perpendicular to the wall. The pressure force exerted on this element per unit of width is $-\rho \frac{dp}{dx} \Delta x$. When the pressure gradient is positive, the pressure force is negative (i.e., is opposite to the flow direction). If the average kinetic energy contained in the boundary layer is not large enough to sustain this opposing force, the flow does not follow the wall any longer and separation occurs (Fig. 16).

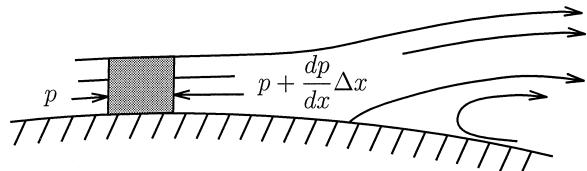


FIGURE 16 Boundary layer separation in a two-dimensional laminar flow.

On an airfoil, separation can occur when the angle of attack increases. On the rear part of the upper surface, the intensity of the positive pressure gradient increases with the angle of attack. Separation is generally not desirable for several reasons. When separation occurs on the rear part of an airfoil, the separated flow is like a dead water zone where the pressure is more or less constant. The pressure distribution on the airfoil is very different from the distribution calculated in an inviscid flow and the airfoil performance is degraded. When the angle of attack increases, the lift decreases and the drag increases because of these pressure changes. The airfoil enters a stall regime. In addition, the flow is likely to become unstable, which is not at all desirable for a transport aircraft.

VII. LAMINAR-TURBULENT TRANSITION

The change from laminar to turbulent regime—the so-called *laminar-turbulent transition*—often originates from an instability phenomenon (Fig. 17). In a two-dimensional flow, the boundary layer is laminar near the leading edge. Downstream, traveling waves occur naturally, and the flow is unsteady. At first, the waves are two dimensional and the corresponding wave vector is parallel to the general direction of the flow. Downstream, complex phenomena occur before the flow becomes turbulent.

The theoretical analysis of the first step of transition is performed using *stability theories*. The results of these theories are relatively well confirmed by experiments. The simplest theory consists of studying the variation of small perturbations. The Navier-Stokes equations are linearized about a basic state—for example, the Blasius boundary layer or any other laminar boundary layer. The longitudinal variations of the basic flow are neglected, and the solutions to the linearized Navier-Stokes equations are sought as normal modes. In a two-dimensional incompressible flow, the continuity equation for the perturbations \hat{u} and \hat{v} of the velocity components is satisfied if a stream function ψ is introduced:

$$\hat{u} = \frac{\partial \psi}{\partial Y}; \quad \hat{v} = -\frac{\partial \psi}{\partial X}$$

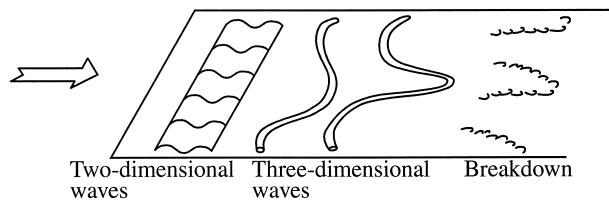


FIGURE 17 Transition process on a flat plate.

The solution for the stream function ψ has the normal mode form:

$$\psi = \phi(Y) e^{-\alpha_i X} e^{i(\alpha_r X - \omega T)}$$

Here, all the quantities are normalized by a reference length ℓ related to the boundary layer thickness and a reference velocity V_0 . The quantity i is the pure imaginary number such that $i^2 = -1$. A perturbation is characterized by its wave number α_r , its pulsation ω , and its amplification coefficient α_i . When $\alpha_i > 0$, the perturbation is damped; when $\alpha_i < 0$, the perturbation is amplified; when $\alpha_i = 0$, the wave is neutral.

The linearized Navier-Stokes equations give the Orr-Sommerfeld equation for the stream function:

$$\begin{aligned} \frac{\partial^4 \phi}{\partial Y^4} - 2\alpha^2 \frac{\partial^2 \phi}{\partial Y^2} + \alpha^4 \phi \\ - iR \left[(\alpha U - \omega) \left(\frac{\partial^2 \phi}{\partial Y^2} - \alpha^2 \phi \right) - \alpha \frac{\partial^2 U}{\partial Y^2} \phi \right] = 0 \end{aligned}$$

where $\alpha = \alpha_r + i\alpha_i$, U is the velocity profile of the basic flow, and R characterizes its Reynolds number:

$$R = \frac{V_0 \ell}{\nu}$$

The boundary conditions of the Orr-Sommerfeld equation express that the velocity perturbations are zero at the wall and vanish at infinity.

In the Orr-Sommerfeld equation, the input is the profile $U(Y)$ and its Reynolds number. A nontrivial solution is obtained for certain combinations of α and ω , which are eigenvalues of the Orr-Sommerfeld equations. Then, for given values of α_r and ω , a nonzero solution is found for certain values of α_i . Therefore, it is possible to say whether a perturbation with a given frequency and wave number is stable or unstable.

The results of this linear theory are used to define *transition criteria*. These criteria are based on the calculation of the amplification rate of the most unstable waves. A transition criterion says that transition occurs when the amplification rate reaches an empirical critical threshold.

The solutions of the Orr-Sommerfeld equation show that the Reynolds number plays an essential role in the stability of a laminar boundary layer and therefore in transition. However, the Reynolds number is a parameter among others. A positive pressure gradient, wall roughnesses, free-stream turbulence, and noise tend to promote transition at a lower Reynolds number. Flow compressibility and wall temperature level are also important parameters.

In a three-dimensional flow on a swept wing (Fig. 18), the secondary flow occurring in the boundary layer can trigger transition when the flow is accelerated, even in the vicinity of the leading edge where a two-dimensional flow would remain laminar. In addition, the turbulent boundary

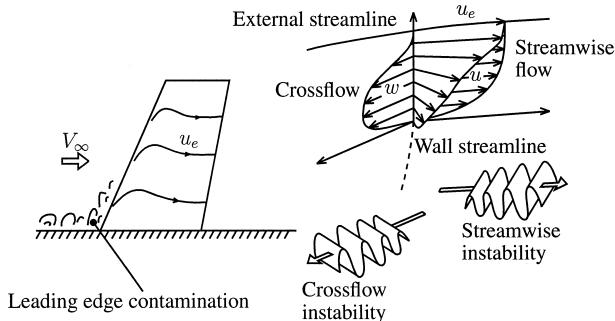


FIGURE 18 Mechanisms of laminar-turbulent transition in a three-dimensional boundary layer on a swept wing.

layer on the fuselage can perturb the boundary layer developing on the leading edge of a swept wing. The leading edge boundary layer is then turbulent and the wing boundary layer is contaminated.

VIII. TURBULENT BOUNDARY LAYERS

A. Effect of Turbulence

An important effect of turbulence is to mix the flow very efficiently. As a result, the inhomogeneities in the flow are smoothed. Compared with a laminar boundary layer, the velocity profile in a turbulent flow is fuller in the core of the flow. At the wall, in a laminar or turbulent flow, the velocity is zero. Then, the velocity profiles resemble those of Fig. 19.

Two conclusions can be drawn from this observation. First, the slope of the velocity profile at the wall is much steeper in a turbulent flow than in a laminar flow. The turbulent skin friction is much higher:

$$\tau_w \text{ turbulent} \gg \tau_w \text{ laminar}$$

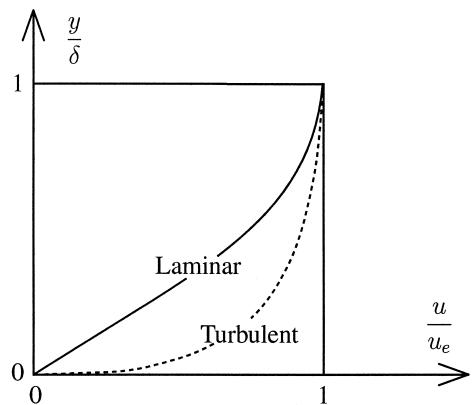


FIGURE 19 Typical velocity profiles in a laminar and a turbulent boundary layer.

Second, the amount of kinetic energy contained in a boundary layer is larger when the flow is turbulent. Separation occurs when the kinetic energy contained in the boundary layer is not sufficient to overcome the pressure force opposing the motion. A turbulent boundary layer is therefore able to sustain a stronger positive pressure gradient than a laminar boundary layer without separating.

B. Turbulent Boundary Layer Structure

A turbulent boundary layer is composed of two layers (Fig. 20):

- Near the wall, the viscous stress is larger than or is of the same order as the apparent turbulent stress. This inner region is studied by introducing the wall variables $y^+ = y \frac{u_\tau}{v}$ and $u^+ = \frac{u}{u_\tau}$. The viscous length scale $\frac{v}{u_\tau}$ is based on the friction velocity $u_\tau = \sqrt{\frac{\tau_w}{\rho}}$ where τ_w is the wall shear stress.
- Away from the wall, the turbulent shear stress is much larger than the viscous stress. This outer layer is fully turbulent and is characterized by the velocity scale u_τ and the length scale δ , the thickness of the boundary layer.

Experiments and theoretical considerations based on the theory of matched asymptotic expansions show that the velocity profile in the inner region follows a universal law $u^+ = f(y^+)$ called the *law of the wall* (Fig. 21). This law is universal in the sense that it does not depend on the conditions under which the boundary layer develops—for example, the value of the Reynolds number or the pressure gradient. In particular, in the part of the inner layer that is far enough from the wall ($y^+ > 50$), it is established that the velocity profile has a logarithmic form:

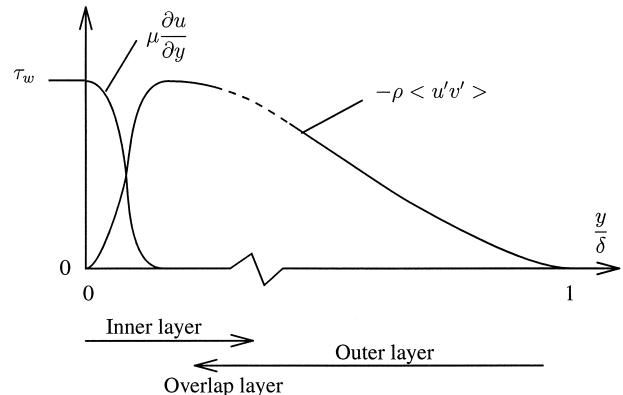


FIGURE 20 Shear stresses in a flat plate turbulent boundary layer. The boundary layer has a two-layer structure. The inner layer thickness is exaggerated in this figure. The inner layer and the outer layer overlap to give a logarithmic velocity law.

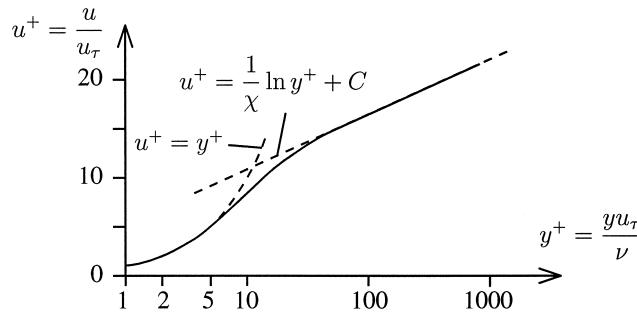


FIGURE 21 Law of the wall in semi-logarithmic coordinates. The logarithmic law is valid for $y^+ < y_L^+$ where y_L^+ depends on the conditions in which the boundary layer develops: Reynolds number and pressure gradient.

$$u^+ = \frac{1}{\chi} \ln y^+ + C \quad (12)$$

where χ is the von Kármán constant. Experimentally, the values of χ and C are $\chi = 0.41$ and $C = 5$. All the turbulence models should reproduce this result in order to model the boundary layer correctly.

C. Simplified Representation of Turbulence

Turbulence models use a description of turbulence based on the concept of scales (velocity, length, time). Two classes of structures play an important role: the *large (energetic) structures* characterized by scales \mathbf{u} , \mathbf{l} , and θ and the *small (dissipative) structures* characterized by ν , η , and τ . The order of magnitude of \mathbf{u} is the square root of the turbulent kinetic energy; the order of \mathbf{l} in a boundary layer is the distance to the wall y for the inner layer and is δ for the outer layer; the time scale is $\theta = \frac{\mathbf{l}}{\mathbf{u}}$.

The dissipation rate ϵ is used to estimate the order of the small scales. By definition, the dissipation rate ϵ represents the amount of turbulent kinetic energy transformed into heat per unit time. The dissipation rate is the deformation work performed by the fluctuations of viscous stresses. For a unit mass, the definition is

$$\epsilon = 2\nu \langle s'_{ij} s'_{ij} \rangle$$

where s'_{ij} is the rate of strain fluctuation.

From this definition, the time scale τ is such that $\epsilon = \frac{\nu}{\tau^2}$. Since the dissipative structures efficiently transform kinetic energy into heat, their characteristic Reynolds number $\frac{\nu \eta}{\nu}$ is on the order of unity. The dissipative or Kolmogorov scales are then evaluated as:

$$\nu = (\nu \epsilon)^{1/4}; \quad \eta = \frac{\nu^{3/4}}{\epsilon^{1/4}}; \quad \tau = \left(\frac{\nu}{\epsilon} \right)^{1/2}$$

An approximate value of ϵ is the flux of energy transferred from the large structures to the dissipative struc-

tures. This flux is governed by inviscid mechanisms, and the amount of dissipation is imposed by the characteristics of the large structures. The dissipation rate is evaluated as:

$$\epsilon = \frac{\mathbf{u}^3}{\mathbf{l}}$$

This result is equivalent to saying that the time scale $\frac{\mathbf{l}}{\mathbf{u}}$ of large structures is equal to the time scale $\frac{\mathbf{u}^2}{\epsilon}$ required to dissipate an amount \mathbf{u}^2 of turbulent kinetic energy.

The ratio of dissipative to energetic scales is

$$\frac{\eta}{l} = R_l^{-3/4}; \quad \frac{\nu}{u} = R_l^{-1/4}; \quad \frac{\tau}{\theta} = R_l^{-1/2}$$

where R_l is a turbulence Reynolds number:

$$R_l = \frac{\nu l}{\nu}$$

As the Reynolds number is much larger than unity, there is a large gap between the small and large scales. The behavior of small-scale structures is nearly independent of the behavior of large-scale structures.

D. Turbulence Models

In a boundary layer, the origin of turbulence is the shear $\frac{\partial u}{\partial y}$. As a general rule, it is assumed that if two phenomena are in efficient interaction, their time scales are of the same order. Therefore, if shear is the only source of turbulence, its time scale is of the same order as the turbulence time scale:

$$\frac{\mathbf{u}}{\mathbf{l}} = \frac{\partial \mathbf{u}}{\partial y}$$

Experiments show that the velocity fluctuations u' and v' are well correlated. The order of $-\langle u' v' \rangle$ is thus \mathbf{u}^2 . The Prandtl mixing length model states:

$$-\langle u' v' \rangle = l^2 \left(\frac{\partial u}{\partial y} \right)^2$$

where the mixing length l represents a turbulence length scale. Near the wall, the length scale of large structures is the distance to the wall $l \sim y$. If it is assumed that $-\rho \langle u' v' \rangle \simeq \tau_w$ (Fig. 20), the logarithmic law is obtained. Precisely, with $l = \chi y$, the logarithmic law (12) is exactly reproduced. Very close to the wall, the model must take into account the effects of the wall on the turbulence structure and modifications are needed. In the outer layer, the length scale is proportional to the boundary layer thickness. Good results are obtained with $l = 0.085\delta$.

The mixing length model was specially developed for wall boundary layers. For wakes or jets, the expression of the mixing length must be modified. More general models use transport equations for turbulence characteristics—for

example, the turbulent kinetic energy k and its dissipation rate ϵ . These equations are deduced from the Navier-Stokes equations written for the velocity fluctuations. The exact equations are not directly a turbulence model because several terms are unknown correlations between velocity fluctuations or between velocity fluctuations and pressure fluctuations. These unknown terms require modeling. The $k - \epsilon$ model is a popular model.

From the inviscid estimate $\epsilon = \frac{u^3}{l}$ and the relation $\frac{u}{l} = \frac{\partial u}{\partial y}$, the shear stress is expressed as:

$$-\langle u'v' \rangle = C_\mu \frac{k^2}{\epsilon} \frac{\partial u}{\partial y}$$

where C_μ is a constant. This expression is the basis of the $k - \epsilon$ model.

The k - and ϵ -model equations for a two-dimensional incompressible shear layer are

$$u \frac{\partial k}{\partial x} + v \frac{\partial k}{\partial y} = -\langle u'v' \rangle \frac{\partial u}{\partial y} - \epsilon + \frac{\partial}{\partial y} \left(C_k \frac{k^2}{\epsilon} \frac{\partial k}{\partial y} \right) \quad (13)$$

$$\begin{aligned} u \frac{\partial \epsilon}{\partial x} + v \frac{\partial \epsilon}{\partial y} &= \frac{\epsilon}{k} \left(-C_{\epsilon_1} \langle u'v' \rangle \frac{\partial u}{\partial y} - C_{\epsilon_2} \epsilon \right) \\ &+ \frac{\partial}{\partial y} \left(C_\epsilon \frac{k^2}{\epsilon} \frac{\partial \epsilon}{\partial y} \right) \end{aligned} \quad (14)$$

where the constants are determined from experimental results. In particular, the properties of the logarithmic layer are used.

The above model is valid for the outer layer of the boundary layer. Near the wall, the model is modified to take into account the wall effects on turbulence.

IX. DRAG REDUCTION

The drag breakdown on various aircraft (business jets of the Falcon type, civil transport aircraft of the Airbus type, supersonic aircraft such as the Concorde) shows three main sources of drag (Fig. 22): friction drag, induced drag, form drag, and parasitic drag.

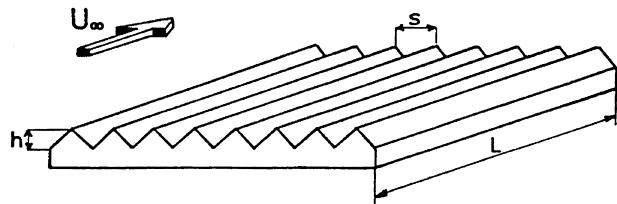


FIGURE 23 Riblet surfaces.

and form drag. *Friction drag* is due to the boundary layer and is obtained by integrating the wall shear stress over the surface of the aircraft. *Induced drag* is associated with the finite dimension of the wings and is due to lift. The pressure is lower on the upper surface of the wings than on the lower surface in order to produce lift. A flow turns around the tip of the wings from the lower surface to the upper surface and generates vortices trailing a long distance behind the aircraft. These vortices modify the effective angle of attack of the wings and turn back the lift vector, leading to a drag component. *Form drag* results from the integration of pressure on the body (excluding induced drag).

In all cases, friction drag accounts for a large share of the total drag of the aircraft and many endeavors have been devoted to reducing this drag.

A. Riblets

Riblet surfaces are made of grooves aligned with the flow (Fig. 23). The role of these surfaces is to reduce the friction drag of a *turbulent* boundary layer.

Experiments in wind tunnels have shown that riblets are efficient in reducing the wall shear stress when the riblet dimensions are adapted to the boundary layer characteristics. The optimal dimensions are expressed in terms of wall units. For triangular shapes with $h = s$, a friction drag reduction of 8% is obtained when $h^+ = s^+$ is in the range of 10–15. In terms of physical dimensions, these figures give a size on the order of a few hundredths of a millimeter. Riblets are efficient in subsonic, transonic, and supersonic flows (Fig. 24). It should be noted that the wetted area (the

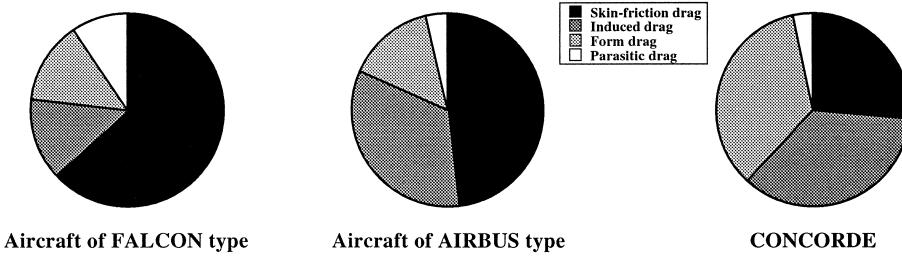


FIGURE 22 Drag breakdown on different types of aircraft.

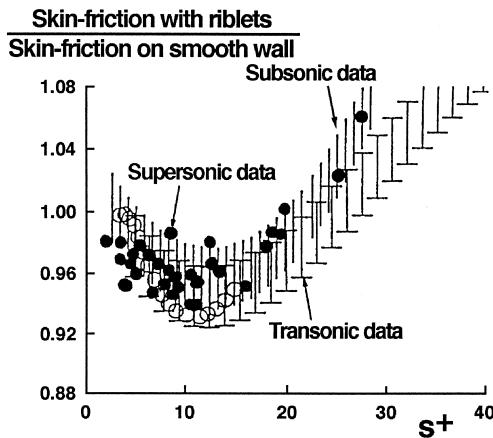


FIGURE 24 Efficiency of riblets.

area in contact with fluid) is much larger for riblet surfaces than for smooth surfaces. Thus, a smooth surface does not produce the lowest friction.

The underlying mechanisms are not completely understood, but it is almost certain that riblets interact with the near-wall turbulent structures which are responsible for turbulence production. The optimal size of riblets is consistent with this explanation.

Flight tests have confirmed the results obtained in wind tunnels. An A320 research aircraft was covered with riblets over 75% of its wetted area. Fuel consumption was measured with and without riblets. For cruise conditions, a reduction of 1.6% of the total drag of the aircraft was recorded. This result is consistent with laboratory results on the friction drag reduction and with the fact that friction drag accounts for about 45% of the total drag of the aircraft and only 75% of the aircraft surface was covered with riblets. In any case, such a drag reduction is significant for an aircraft.

B. Laminarity

An even more efficient means of reducing friction drag is to *laminarize the boundary layer*, as skin friction is much lower in a laminar flow than in a turbulent flow. It is more relevant to laminarize the upper surface because the turbulent skin friction is higher than on the lower surface.

Laminarization is commonly used for sailplanes. The airfoil shape is designed to generate a negative pressure gradient over a large area of the wing surface. On a sailplane, the wing has no sweep and the flow on the wings is nearly two dimensional. A negative pressure gradient has the property of stabilizing the laminar regime. A typical pressure distribution on the upper surface of a wing as shown in Fig. 25a is conducive to the laminar regime in two-dimensional flow over a significant extent.

In a three-dimensional flow, transition can occur due to the instability of the boundary layer streamwise flow or of the crossflow (Fig. 18). The streamwise instability occurs under the same conditions as in a two-dimensional boundary layer. A negative pressure gradient increases the transition Reynolds number, whereas a positive pressure gradient reduces it. However, crossflow instability can develop in the presence of an accelerated flow, whereas the streamwise flow is stable.

Wing shaping is more complicated for preserving natural laminarity in a three-dimensional flow. For moderate sweep angles (less than 20°), a typical pressure distribution is shown in Fig. 25b for the upper surface of a wing. Such a pressure distribution results from a compromise designed to minimize streamwise and crossflow instabilities. Near the leading edge, the flow is strongly accelerated; in this way, the streamwise instability is eliminated and the crossflow instability is delayed. Downstream, a milder acceleration minimizes the risk of streamwise instability and

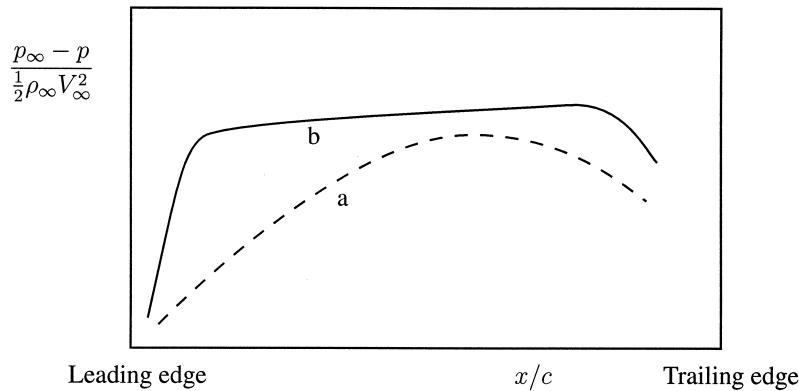


FIGURE 25 Pressure distribution producing laminar boundary layers (a) in a two-dimensional flow, and (b) on a swept wing.

the crossflow development is limited because the pressure gradient is moderate.

For higher Reynolds numbers or for higher sweep angles, wing shaping is not sufficient to maintain a laminar regime. Another technique consists of applying *boundary layer suction* at the wall. In practice, suction is performed through perforated walls. Very tiny holes are drilled by means of laser techniques, for example; the holes are a few hundredths of a millimeter in size and are spaced a few tenths of a millimeter apart. Wall suction has the property of decreasing the boundary layer thickness and modifying the shape of the velocity profile, which is more stable. Wall suction reduces the intensity of the cross-flow which is very favorable to maintaining the laminar regime.

Leading edge contamination is a third mechanism that can prevent the boundary layer from being laminar (Fig. 18). Leading edge contamination is characterized by a Reynolds number \bar{R} which increases when the sweep angle ϕ increases, when the free-stream velocity V_∞ increases, and when the leading edge radius r_0 increases. For a swept circular cylinder, this Reynolds number is

$$\bar{R} = \left[\frac{V_\infty r_0 \sin \phi \tan \phi}{\nu} \frac{1}{2} \right]^{1/2}$$

From experiments, the boundary layer is laminar when $\bar{R} < 250$; however, for a civil transport aircraft, \bar{R} is above this critical value, at least near the wing root.

Two approaches can be employed to delay leading edge contamination. One possible approach is to use a so-called Gaster bump—a device proposed by Gaster—which is a small bump placed on the wing leading edge at a certain distance from the fuselage, well outside the fuselage boundary layer. The bump diverts the flow coming from the fuselage and creates a stagnation point from which a fresh boundary layer develops. This device can delay the onset of leading edge contamination up to values of \bar{R} around 350 or 400. A much more efficient way of delaying leading edge contamination is to use suction in the wing leading edge region. With relatively moderate suction rates, leading edge contamination has been delayed up to $\bar{R} = 700$, but this value is certainly not the maximum value that can be reached with wall suction.

Flight tests have confirmed the possibility of laminarity on a civil transport aircraft or on a business jet. For example, experiments were performed on a Boeing 757 which flew with wall suction on the wing. The boundary layer was laminar up to the shock wave (i.e., about 60% of the chord). In Europe, much effort has been devoted to the study of laminarity, and flight tests were performed on a Fokker 100 to analyze natural transitions. In France, flight tests were performed on different Falcon aircraft. In par-

ticular, the two wings of a Falcon 900 were equipped with wall suction and this demonstrator was conceived in an operational framework (suction system, anti-icing system providing protection against insect pollution). An A320 laboratory aircraft was also tested in flight with a laminar fin equipped with wall suction. Laminarity on a vertical fin is relatively simple because a fin has no high lift device and integration of the suction system is easier. A fin also has the advantage of being symmetrical. However, the high angle of sweep is a real challenge for laminarity. Flight tests were very successful, as laminarity was obtained over 60% of the chord.

X. CONCLUDING REMARKS

The boundary layer concept was invented at the beginning of the twentieth century (i.e., about hundred years ago). At that time, this very inspired approximation of Navier-Stokes equations appeared as a breakthrough in aerodynamics and opened the doors to new knowledge. It has been said that the boundary layer was the key to aerodynamics.

Since then, intensive efforts have been devoted to this field and the progress made is really impressive. For example, the problem of attached laminar boundary layers can be considered as solved. The question of boundary layer separation is now well understood and, although this problem has not been completely solved, practical solutions exist.

In addition, numerical solvers of averaged Navier-Stokes equations are becoming a routine tool in research laboratories and in aircraft companies. At least in principle, these solutions are able to describe the complete flow field around an aircraft. The question, then, is whether the boundary layer concept is still valuable or is old-fashioned. Undoubtedly, the answer is that the boundary layer remains an invaluable model. At least two reasons can be given. First, understanding the behavior of the flow around a wing, for example, requires understanding the boundary layer. The correct interpretation of numerical or experimental results often requires a good grasp of viscosity effects. Second, many questions are still unsolved. Among them, laminar-turbulent transition and the turbulent regime are two of the most difficult problems in physics. Boundary layers, wakes, and jets are obviously places where turbulence plays an essential role. The understanding of turbulence development is far from being complete and its modeling is not at all satisfactory. Clearly, work is needed in this field and wall boundary layers are an appropriate model for these studies.

Another challenging topic is drag reduction. Boundary layers have an important contribution to drag. As mentioned in a previous section, different techniques are

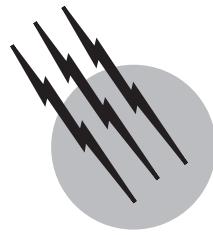
available to reduce skin-friction drag significantly. Active control of turbulence is another way that is under study at the present time and for which new ideas are emerging continuously. The principle of active control is to detect the development of turbulence from the measurement of some characteristic feature—wall pressure or wall shear stress, for example—and to apply a perturbation to the flow in order to cancel this development. Up to now, these studies have been performed in the laboratory, either experimentally or numerically. In other fields, active control techniques are used in real life. For example, noise reduction is already achieved by using such techniques. It is reasonable to expect that turbulence manipulation will become a reality.

SEE ALSO THE FOLLOWING ARTICLES

AIRCRAFT PERFORMANCE AND DESIGN • AIRCRAFT SPEED AND ALTITUDE • AIRPLANES, LIGHT • COMPUTATIONAL AERODYNAMICS • FLIGHT (AERODYNAMICS) • FLOW VISUALIZATION • FLUID DYNAMICS

BIBLIOGRAPHY

- Cebeci, T., and Cousteix, J. (1999). "Modeling and Computation of Boundary-Layer Flows," Springer-Verlag, Berlin/New York, and Horizons Publishing, Long Beach, CA.
- Drazin, P. G., and Reid, W. H. (1981). "Hydrodynamic Stability," Cambridge University Press, Cambridge, U.K.
- Lesieur, M. (1990). "Turbulence in Fluids," Kluwer Academic, Dordrecht/Norwell, MA.
- Piquet, J. (1999). "Turbulent Flows: Models and Physics," Springer-Verlag, Berlin/New York.
- Sagaut, P. (1998). "Introduction à la simulation des grandes échelles pour les écoulements de fluide incompressible," Springer-Verlag, Berlin/New York.
- Schlichting, H. (1968). "Boundary-Layer Theory," McGraw-Hill, New York.
- Tennekes, H., and Lumley, J. L., (1972). "A First Course in Turbulence," MIT Press, Cambridge, MA.



Aircraft Instruments

W. B. Ribbens

University of Michigan

- I. Introduction
- II. Aircraft Communication Systems
- III. Navigation
- IV. Distance Measuring Equipment
- V. LORAN
- VI. Automatic Direction Finding
- VII. Long-Range Navigation
- VIII. Navigational Coordinate Systems
- IX. Satellite-Based Navigation System
- X. The GPS System Structure
- XI. WAAS

- XII. GPS Accuracy Augmentation
- XIII. GPS Integrity and Availability
- XIV. Receiver Autonomous Integrity Monitoring (RAIM)
- XV. Inertial Navigation
- XVI. Instrument Landing System (ILS)
- XVII. Additional Precision Approach Systems
- XVIII. Air Data System
- XIX. Attitude and Heading References
- XX. Autopilot—Flight Management Systems (FMS)
- XXI. Glass Cockpit

GLOSSARY

Aircraft instrumentation Electronic apparatus necessary for voice/digital communication, navigation, attitude measurement, flight path control, and system monitoring.

Air data system Instrumentation for accurately calculating critical flight variables (e.g., airspeed, altitude) from onboard measurements.

Altitude and heading instrumentation Electronic equipment for measuring and displaying to the flight crew the aircraft orientation relative to Earth coordinates and its velocity vector components.

Flight management system Electronic system for automatically controlling aircraft flight path along segments of its route from departure to destination.

Fly by wire All electronic flight control system for con-

trolling attitude having the feature of variable aircraft stability.

Full authority digital engine control (FADEC) Electronic engine control system for regulating engine performance within a safe operating envelope with command inputs from the flight crew.

Glass cockpit Solid state display system capable of alphanumeric data or pictorial display to the flight crew eliminating the need for any electromechanical display.

Inertial navigation system (INS) Electronic position/velocity vector determination derived from measurements of acceleration along and about aircraft axes.

Instrument landing system Instrumentation for precisely and accurately measuring and displaying to the flight crew the aircraft flight path during the landing phase of flight used under severely adverse weather conditions.

Radio navigation Aircraft location/position determination based upon signals from ground-located transmitters.

Satellite navigation Position/location and velocity vector determination derived from signals transmitted from a constellation of satellites (global positioning system, GPS).

Transceiver Two-way radio voice communication between flight crew and controlling authority.

I. INTRODUCTION

Aircraft instrumentation provides the flight crew with the capability to perform many functions, including communication, navigation, monitoring the status of onboard systems, diagnosing problems or system failures that occur in flight, measuring/monitoring aircraft attitude, monitoring weather, and monitoring the position of other aircraft (i.e., those that pose a potential collision hazard). The actual instrumentation required depends upon the aircraft category and the type of flying involved.

The least capability required of instruments is for light, single-engine, general aviation aircraft licensed by the Federal Aviation Administration (FAA) under Federal Air Regulations (FAR) part 23 for day visual flight rules (VFR) operations. Instrumentation progresses from this category and type of flying to greater and greater complexity, reaching the greatest requirement for transport category aircraft (licensed by the FAA under FAR part 25) and flying under instrument flight rules (IFR). The appropriate FARs for the first category are under FAR part 61. For noncommercial instrument flying FAR part 91 specifies requirements. Commercial operations other than regularly scheduled airline operation (including air taxi and air cargo operations) are conducted under FAR part 135 rules. For airline operations, the rules are given in FAR part 121. Each aircraft category and flight operation has instrumentation requirements such that the higher category requirements tend to be a superset of requirements for lower requirements. It is perhaps instructive to begin with the simplest aircraft instrumentation and progress through the various categories and flight operations to the most complex.

The minimum instrumentation required for day VFR flying in light aircraft includes the following:

1. Airspeed indicator
2. Altimeter
3. Magnetic direction finder (compass)
4. Fuel gauge indicating quantity of fuel in each tank
5. Landing gear position indicating device for retractable landing gear aircraft

6. Various engine performance and operating safety parameters, including
 - a. Oil pressure and temperature
 - b. Angular speed of critical rotating parts (e.g., crankshaft)
 - c. Airflow monitoring gauge (e.g., manifold pressure)

For operations under IFR under FAR part 91, additional equipment beyond that required for VFR operation includes the following:

1. Radio equipment suitable for navigation using navigational facilities to be used during the flight as well as two-way voice and/or data communications between the aircraft and the controlling agency
2. Aircraft attitude indicating equipment for displaying pitch, roll, and yaw
3. Slip/skid indicating equipment
4. The altimeter which is adjustable for local atmospheric pressure
5. Precision clock (HH = hours, MM = minutes, SS = seconds)
6. Aircraft heading and turn rate indicating equipment

It should be noted that in traditional light (general aviation) aircraft, some of this equipment was electronic (i.e., avionics) and some was mechanical.

The avionics suite installed on any given airplane is a function of the aircraft category (i.e., general aviation or transport etc.) and upon its intended operations (e.g., part 91 or 135 or 121). Presented below is a listing of avionics equipment that is a superset of the possible choices for avionics equipment. The listing below is based upon functionality of the equipment.

1. Communications
 - a. Ground-based voice—air traffic control (ATC)
 - b. Ground-based data link (ATC future)
 - c. Satellite based
2. Navigation
 - a. Radio land based
 - b. Satellite
 - c. Inertial platform
3. Landing aids
 - a. Land based
 - b. Satellite based
4. Attitude indicating
5. Air data system
6. Flight control
7. Flight safety

II. AIRCRAFT COMMUNICATION SYSTEMS

Aircraft communication systems function for the sole purpose of providing a communication link between each individual airplane and the appropriate ATC authority. The appropriate control authority is a function of the phase of the flight and the nature of the flight (i.e., VFR or IFR) but is independent of the aircraft category, that is, to say all civil aircraft have similar (in function) radio communication equipment from the smallest general aviation aircraft through the largest part 121 transport aircraft. Military aircraft have additional radio equipment beyond that which is required for flying in civilian airspace. However, this equipment differs only in certain detailed aspects (e.g., carrier frequency band) from its civilian counterpart.

There are several levels of control authority with which the flight crew must establish two-way radio communication. At the larger airports serving air carriers, communications typically begin with communication with "clearance delivery." For IFR operations, the crew will have previously filed a flight plan with the ATC system. (Note: this can be filed verbally via telephone to a government agency known as a flight service station (FSS) for private general aviation or via a computer link for an air carrier.)

The clearance delivery authority can verify the details of the flight plan and its acceptance by ATC. Any changes that might be required (e.g., due to a radio navigation facility being temporarily out of service) can be made during the exchange with a branch of ATC known as clearance delivery. At the end of the communication with clearance delivery, the flight plan has been finalized and accepted by both ATC and the flight crew.

The next level of authority contacted by the crew is the so-called ground control which controls aircraft ground movement. The flight crew will maintain two-way radio communication via a specific carrier frequency (e.g., 121.7 MHz) with ground control until the aircraft has taxied to the active runway.

At this point the crew is instructed to contact the control tower on its assigned carrier frequency. The control tower personnel have the authority to regulate aircraft movement on the ground and in the air in the vicinity of the airport (typically within a 5-mile radius of the airport). This authority issues clearance to taxi onto the runway and then, when it is safe, to take off. Two-way radio communication with the control tower is maintained by the crew until they are specifically instructed to contact departure control on a specific carrier frequency.

The departure control authority regulates aircraft movement in predetermined specific airspace in the vicinity of the airport. Departure control personnel have access to

surveillance radar data that can be used to maintain safe aircraft separation. Instructions are given to the crew to fly specific trajectories via assigned heading, airspeed, and altitude.

At some appropriate point in the airspace, the crew will be instructed to contact the next controlling authority on a specified carrier frequency. For flights conducted above a certain altitude (e.g., 8000 ft), the controlling authority will be one of the ATC centers. The ATC centers are at widely spaced locations, e.g., Cleveland, Chicago, Minneapolis, etc. This process of "hand-off" from one control authority to the next continues until the aircraft reaches its destination, where it will sequentially maintain two-way radio communication with approach control, tower, and ground control authorities.

The communication radios in civilian aircraft are essentially conventional transceivers that operate in the VHF portion of the spectrum. The carrier frequency f_c is in the band

$$118.000 \leq f_c \leq 136.975 \text{ MHz}$$

in 25 kHz increments offering 760 separate channels. Each controlling authority is assigned a specific carrier frequency along with the portion of airspace that it must control. Normally, dual aircraft communication transceivers are fitted to the aircraft for redundancy.

In modern aircraft communication systems, the tuning for both transmitter and receiver is crystal controlled. A digital frequency synthesizer generates signals at each carrier frequency within the VHF aircraft radio band (on all 760 channels). [Figure 1](#) is a simplified block diagram of a typical aircraft communication transceiver.

The aircraft communication transceiver functions together with a separate electronic module called an "audio panel." This device routes the microphone input from the crew microphones to the particular transmitter being used and routes audio output from the particular receiver in use to the crew headphones/loudspeaker.

The transmitter is a conventional amplitude-modulated (AM) system in which the amplified audio signal from the crew microphones varies the carrier amplitude. Let the carrier frequency selected by the pilot be f_c . The output of the digital frequency synthesizer is $e_{fs}(t)$:

$$e_{fs}(t) = E_{fs} \sin(2\pi f_c t).$$

For each sinusoidal component in the voice (audio) signal at a frequency, f_m , the signal from the modulator, e_{AM} , is given by

$$e_{AM}(t) = E_{AM}[1 + m \sin(2\pi f_m t + \phi)] \sin(2\pi f_c t),$$

where m (the modulation index) is proportional to the amplitude of the audio signal and $m < 1$.

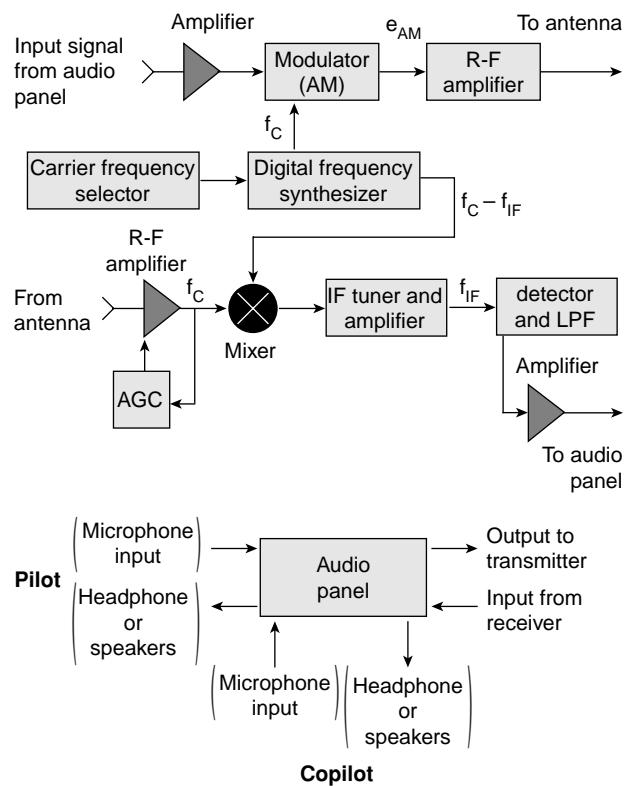


FIGURE 1 Communication transceiver block diagram.

The receiver portion of this system is a conventional superheterodyne receiver. The input signal from the aircraft antenna is proportional to the transmitter signal ($e_{AM}(t)$). This signal is amplified and sent to the mixer. The mixer is a two-input electronic circuit that is functionally equivalent to an analog multiplier or similar nonlinear function of its two inputs. The other input to the mixer is a sinusoid at frequency $f_c - f_{IF}$, where f_{IF} is an intermediate frequency (IF). The output of the mixer includes components at the sum and difference of its inputs (i.e., $f_c \pm (f_c - f_{IF})$), both of which have the same modulation as the input signal from the antenna.

The intermediate frequency amplifier is a narrow band filter/amplifier tuned to f_{IF} such that its output is proportional to

$$[1 + m \sin(2\pi f_m t + \phi)] \sin(2\pi f_{IF} t).$$

This signal is an AM signal at a fixed frequency (f_{IF}). The detector/low pass filter (LPF) demodulates the audio signal and sends the voice (audio) to the audio panel for distribution to the crew headsets/speakers.

At major airports during certain periods of the day, the controlling authorities are exceptionally busy. The controller is often continuously transmitting instructions, pausing between successive calls to wait for a pilot reply

(the crew repeats ATC instructions to confirm that the message was received and acknowledged). In an attempt to reduce the workload burden and to improve system efficiency, experiments are being conducted in which certain instructions are relayed to the flight crew through a digital data link. It is anticipated that the digital data link will provide a supplement to the voice link rather than a replacement for it.

III. NAVIGATION

Except for relatively low altitude (e.g., $h \leq 8000$ ft above ground level) and/or local area flying, all aircraft require navigation aids. In principle, under VFR conditions, point to point navigation can be accomplished with acceptable accuracy to locate the airport of intended landing by means of a suitable map, a magnetic compass, and a clock. In practice, and particularly under IFR conditions, greater accuracy and precision in navigation are required than that available from the above.

Electronic navigation aids, coupled with attitude measuring equipment, have provided sufficient precision and accuracy to conduct poor weather operations since the early 1930s. However, significant progress in the performance of such navigation aids has evolved in the intervening years. It is beyond the scope of this article to provide a history of that evolution. Rather, a survey of the technology as of the late 20th century and the early 21st century is provided specifically with reference to IFR flight conditions.

The electronic navigation aids utilized in any flight depend upon the phase of the flight and the actual weather conditions. The greatest precision and accuracy are required during the final phase of landing an aircraft in the lowest visibility weather conditions. The technology exists to provide navigation with sufficient accuracy so that when coupled to a flight control system the aircraft can complete an automatic landing under zero-zero conditions (i.e., ceiling of zero and zero visibility) using what is known as a category IIIC landing system. However, to understand this system, it is helpful to review navigation aids of lower precision and accuracy.

The navigation on a point to point flight is conducted via preplanned routes known as airways. These airways consist of segments of essentially locally straight lines or segments of great circle areas connecting radio navigation aids and or points in space that can be located by means of radio navigation aids. The accuracy and precision of on-board navigation aids must be such as to permit flight within 4 nmiles of the centerline of the airway. This is a significantly lower requirement than the accuracy required of airport landing navigation aids.

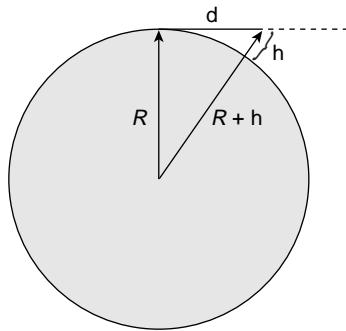


FIGURE 2 Illustration of radio reception range vs altitude.

Navigation along the airways can either be by means of radio-based or inertial platform-based equipment. We consider radio-based equipment first. These, in turn, can achieve navigational measurements from land-based or satellite-based radar transmitters.

The land-based, cross-country (as opposed to landing) radio-navigation system with the greatest accuracy utilizes transmitters operating with carrier frequencies (f_c) in the very high frequency VHF band from $108 \leq f_c \leq 118.0$ MHz. This system is known as the “Very high frequency—omni Range (VOR) navigation system.”

As this system operates with a VHF carrier frequency, its propagation is essentially line of sight. Aside from receiver noise limitations, the range of the VOR system is a function of aircraft altitude in the form of the distance from the aircraft to the local horizon. It can be seen with reference to Fig. 2 that the range d for an aircraft operating at altitude h above the Earth (radius R) is given by

$$d = (2hR + h^2)^{1/2} \cong \sqrt{2hR},$$

where the approximation is valid since $R \gg h$. IFR flight is conducted at altitudes of 3000 ft or greater such that the effective range is more than 65 nmiles.

The VOR system in North America and Europe consists of a set of radio transmitters that are spaced in an irregular grid covering the entire landmass. The average spacing between VOR stations is about 60–80 nmiles (depending upon geographical region), permitting successful VOR navigation between any airport pair on each of these continents.

The VOR navigation system provides a means of locating an aircraft position relative to a given VOR station. This position is specified by two vectors. One vector projects from the station location and is directed radially away from the station. The angle measured clockwise (looking down on the station) from magnetic north is known as the radial (R). The VOR receiver in the aircraft yields a measurement of the radial along which the aircraft is located.

The other vector is directed from the aircraft location toward the VOR station. The angle of this vector measured clockwise (as viewed from above the aircraft) from magnetic north is known as the bearing (B) to the station. These angles are related by

$$B = R + 180^\circ (\text{modulo } 360^\circ).$$

A pair of vectors is also defined relative to the aircraft. The first of these is a vector pointed forward along the longitudinal axis of the aircraft. The angle measured from magnetic north clockwise (looking down) is known as the heading (H). Another vector is the ground velocity vector. The angle measured clockwise from magnetic north is known as the ground track angle or track (T), and the magnitude of this vector is known as ground speed and is the speed of the aircraft relative to an Earth fixed coordinate system. The intended value for this angle is known as the course. Heading and track angles are generally different except in a zero wind zero sideslip condition. It should be noted that the VOR radial and bearing are independent of bearing, track, or course.

The VOR navigation equipment can locate an aircraft along a given radial but does not, by itself, give the distance yielding the aircraft two-dimensional (2-D) position or map coordinates. This 2-D position can be found by using any two VOR stations that are not colinear with the aircraft position (as illustrated in Fig. 3) by triangulation.

An alternate method of measuring the aircraft 2-D position is possible with some VOR stations. These stations, which are designated VOR-DME stations, incorporate distance measuring equipment (DME) colocated with the VOR station. Of course, the aircraft must also be equipped with the corresponding DME equipment. (DME is explained later in this section.)

The operation of the VOR transmitting station can be understood with reference to the functional block diagram of Fig. 4.

The instantaneous frequency $f_{FM}(t)$ is given by

$$f_{FM} = 9960 + 480 \sin(60\pi t).$$

That is, f_{FM} varies from 9480 to 10,440 Hz at a 30-Hz rate. This frequency-modulated (constant amplitude) subcarrier amplitude modulates the carrier yielding

$$\{1 + m \sin[\phi_F(t)]\} \sin(2\pi f_c t),$$

where

$$\phi_F(t) = 2\pi \int_0^t f_{FM}(\tau) d\tau = \text{total phase of subcarrier.}$$

The operation of the VOR system is strongly dependent upon the transmitting antenna system. This antenna has a

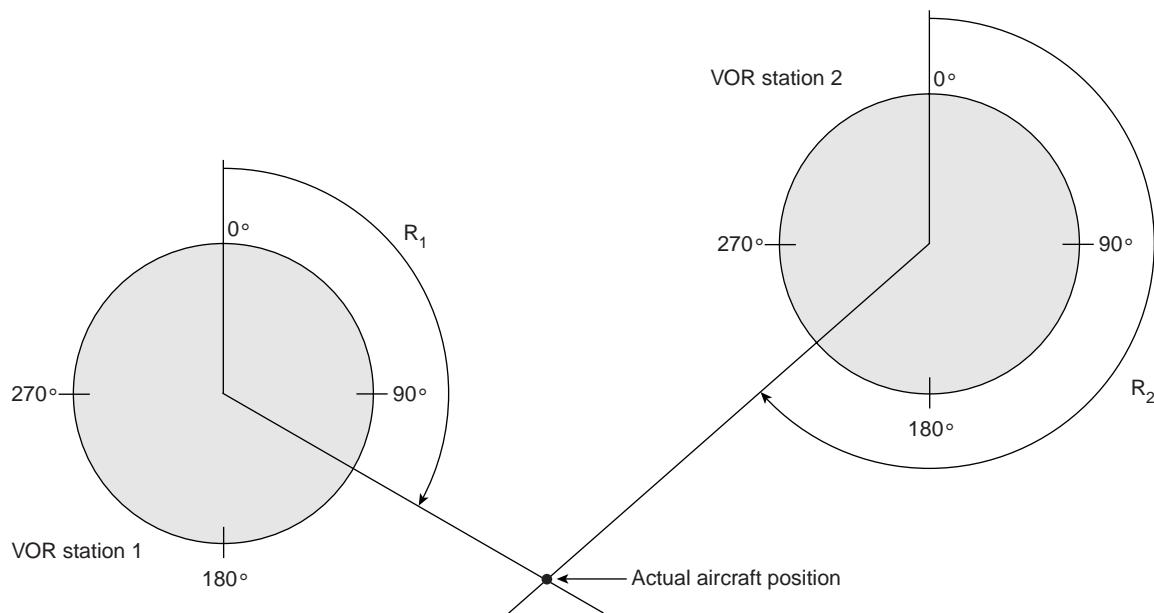


FIGURE 3 Illustration of navigation by dual VOR stations.

unique directional pattern in which the signal strength E varies with angle θ in the form

$$E(\theta) = E_0(1 + \cos \theta).$$

The antenna system causes this pattern to rotate (electronically) at 30 revolutions per second. The effect of this system is functionally equivalent to an antenna with a spatially fixed pattern $E(\theta)$ that is mechanically rotated at 30 revolutions per second. The result is that at any given location the VOR signal is amplitude modulated at 30 Hz.

The operation of the VOR receiver can now be understood with reference to [Fig. 5](#).

The output of the r-f tuner and amplifier is an AM carrier (at the carrier frequency to which the receiver is tuned). This signal is amplitude modulated at 30 Hz due to the rotating antenna radiation pattern and at the 9960-Hz sub-

carrier. These two signals are separated by a pair of narrow band, bandpass filters at center frequencies 30 Hz (yielding e_1) and 9960 Hz.

It should be noted that the phase of the 30-Hz signal e_1 is a function of the radial at the aircraft position. The 30-Hz signal e_2 has a fixed phase that is chosen such that the signals e_1 and e_2 are exactly in phase when the aircraft is along a 0° radial [i.e., directly north (magnetic) of the VOR station]. Denoting ϕ as the phase angle of e_2 relative to e_1 ,

$$\phi = \angle e_1, e_2 = \theta,$$

that is, this phase is equal to the radial along which the aircraft is located.

Navigational position measurements (θ) are made by measuring phase angle ϕ . The measurement is accomplished in conjunction with a special purpose display and a variable phase shifter. Although there are multiple practical implementations of this display function, we will

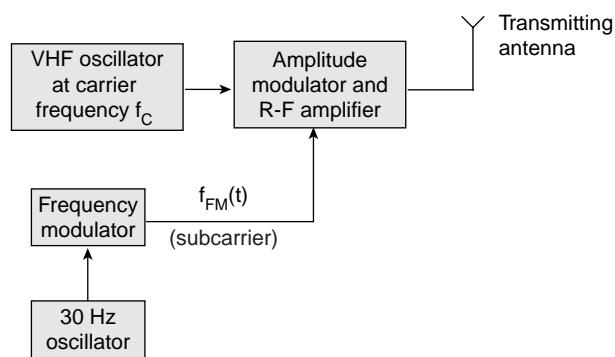


FIGURE 4 Functional block diagram of the VOR transmitter.

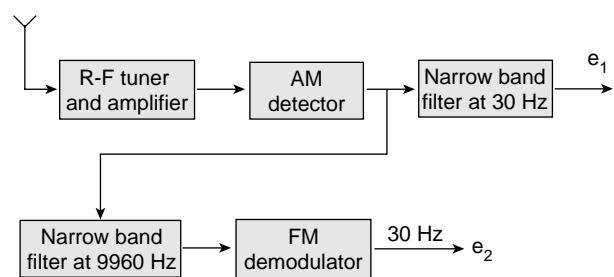


FIGURE 5 VOR receiver block diagram.

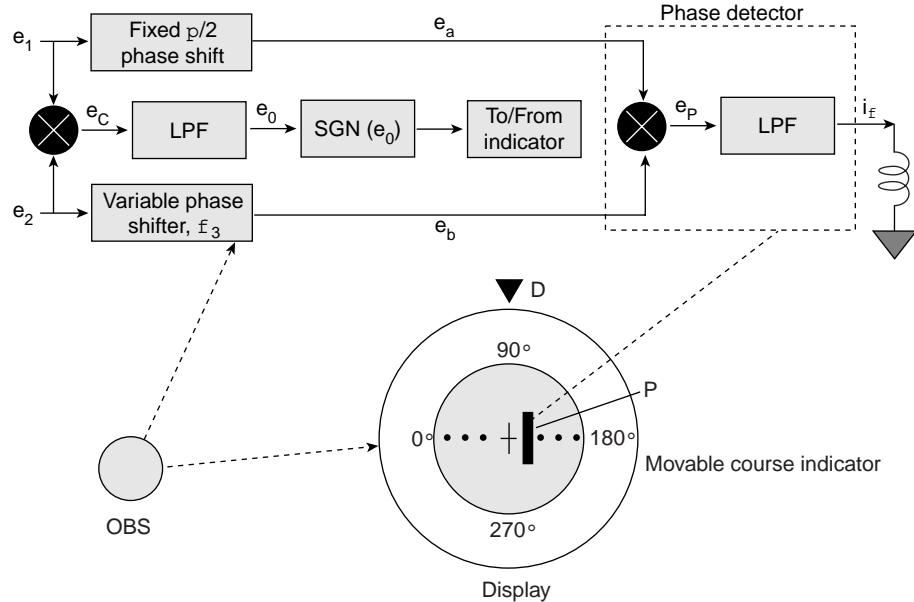


FIGURE 6 Functional block diagram of a phase angle measurement system.

illustrate the general principles involved, with a specific configuration as depicted in Fig. 6 that is, at least, representative of the functional operation of such a subsystem.

In Fig. 6, a hypothetical phase detector is shown consisting of an analog multiplier (mult) and a low pass filter (LPF). Choosing the phase of e_1 as the phase reference, we have

$$e_1(t) = \sin 60\pi t$$

$$E_2(t) = e_2 \sin(60\pi t + \phi)$$

and

$$e_a = e_1 \sin\left(60\pi t + \frac{\pi}{2}\right)$$

$$e_b = e_2 \sin(60\pi t - \phi + \phi_s)$$

$$e_p = ke_1 e_2 \left[\sin\left(60\pi t + \frac{\pi}{2}\right) \right] [\sin(60\pi t - \phi + \phi_s)],$$

where ϕ is the phase angle introduced by the variable phase shifter,

$$= \frac{k}{2} e_1 e_2 [\cos(120\pi t - \phi + \phi_s) + \sin(\phi - \phi_s)]$$

The LPF suppresses the 120-Hz component and generates current e_ϕ :

$$e_\phi = K \sin(\phi - \phi_s),$$

This steady current passes through the coil of a galvanometer whose movable element deflects pointer P on the face of the display. The variable phase shifter is mechanically linked to a knob known as the OMNI Bearing Selector (OBS) that is further linked to a movable course

indicator calibrated for ϕ_s in degrees from 0 through 360° (module 360°). This OBS is adjusted by the flight crew until the pointer P (also known as the course deviation indicator, CDI) is centered. This occurs for $i_\phi = 0$, which occurs for $\phi_s = \phi$. Since $\phi = \theta = \phi_s$, the phase shift introduced by the variable phase shifter when the needle is centered yields a measurement of the radial or bearing of the aircraft. This value is numerically indicated by the movable course indicator under pointer D. The pilot can then read position from the point on the movable course indicator directly beneath the marker D.

As long as the aircraft flies along the radial corresponding to $\phi = \phi_s = \theta$, the meter will remain centered. However, as the aircraft position deviates from the selected radial, the CDI deflects proportionately. The dots on either side of center correspond to 5° deviations for each dot. This display reaches saturation for deviation of $\pm 12^\circ$ from the selected radial.

There is an ambiguity in the radial corresponding to aircraft position of 180°, since there are two nulls for ϕ_s at $\phi_s = \phi$ and at $\phi_s = \phi + 180^\circ$. This ambiguity is resolved by a binary display indicating TO/FROM. As long as the aircraft is located such that

$$\phi_s - \pi/2 < \phi \leq \phi_s + \pi/2,$$

the indicator displays FROM. Otherwise, it displays TO. A scheme for achieving the TO/FROM is depicted in Fig. 6. The two signals from the VOR receiver are functionally multiplied and low pass filtered. The sign of the result drives the binary TO/FROM indicator.

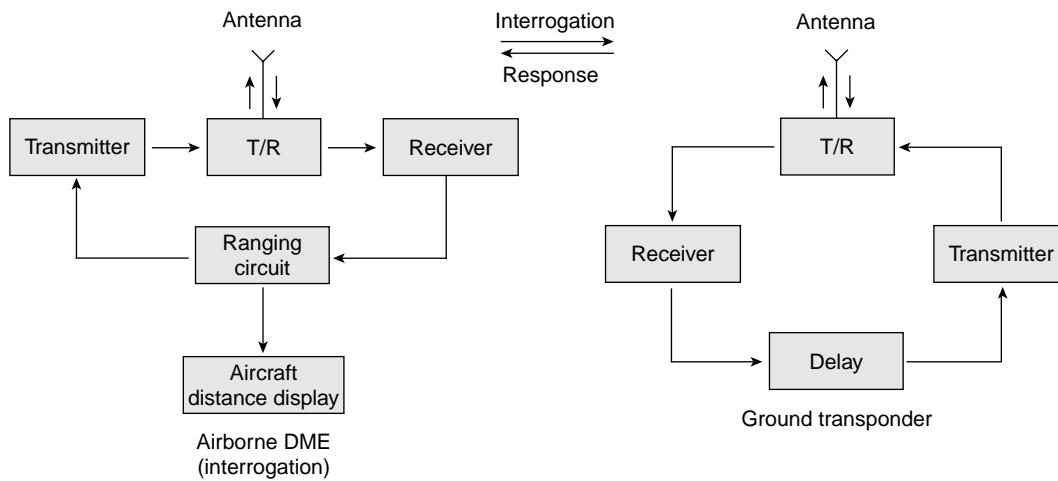


FIGURE 7 Block diagram of a DME system.

IV. DISTANCE MEASURING EQUIPMENT

Distance measuring equipment (DME) is located with VOR ground stations and provides a measurement of the distance of the aircraft user from the associated VOR station. A combination of VOR/DME measurements yields the vector position of the aircraft relative to the VOR/DME radio station. The DME concept is based upon the propagation time of pulses transmitted by the aircraft to a transponder repeater located at the ground station and on the known propagation speed of these pulses. A block diagram of the system is depicted in Fig. 7.

The aircraft interrogation equipment transmits a pair of pulses on one of 126 carrier frequencies spaced 1 MHz apart in the range $1025 \leq f_c \leq 1150$ MHz. The pulse pairs are $12 \mu\text{s}$ apart and have a duration of $3.5 \mu\text{s}$. The pulse pair repetition frequency is between 5 and 150 per second and the peak power is between 50 W and 2 kW depending upon the equipment design. The ground equipment receives these pulses, introduces a fixed $50-\mu\text{s}$ delay, and then retransmits them. The retransmitted pulses are on a carrier frequency $f_{cr} = f_c \pm 63$ MHz.

The aircraft equipment receives these pulses and measures the time delay Δt from transmission to reception and converts the time delay to a distance measurement D :

$$D = c(\Delta t - 50 \mu\text{s})/2,$$

where c = speed of propagation of the modulated carrier wave.

Each ground transponder is designed to handle at least 50 but typically about 100 aircraft. The pulse repetition rate is intentionally made variable, and the interrogator is designed to respond only to retransmitted pulses whose pulse-repetition rate and phase are exactly those transmitted.

V. LORAN

Another radio-based navigation aid that is used particularly by general aviation aircraft is known as the Long-Range Navigation system or LORAN. LORAN is a hyperbolic navigation system, so-called because it locates an aircraft along a hyperbola whose foci are a pair of radio transmitter stations. The intersection of hyperbolas from multiple pairs of stations determines the aircraft position along the Earth's surface (i.e., yields the solution to the navigation problem).

The current version of LORAN evolved from a system developed in Europe before World War II and is known as LORAN-C. LORAN-C consists of a system of 24 transmitter stations in North America and Russia that operate in groups forming so-called "chains." Each chain consists of at least three stations (although four is more common). One of these in each chain is a master and the others are secondaries. Transmitter power level P is in the range ($400 < P < 1000$ kW).

Each chain transmits pulses of the very specific format of a 100-kHz carrier frequency f_c having amplitude such that the antenna current $I(t)$ is given by

$$I(t) = \begin{cases} A(t - \tau)^2 \exp\left\{\frac{-2(t - \tau)}{T}\right\} \sin(2\pi f_c t + \phi) & t > \tau \\ 0 & t < \tau, \end{cases}$$

where t = clock time

τ = envelope to cycle difference

$T = 65 \mu\text{s}$

ϕ = phase code parameter (radians)

$\phi = 0$ for positive phase

$= \pi$ for negative phase

A = amplitude parameter

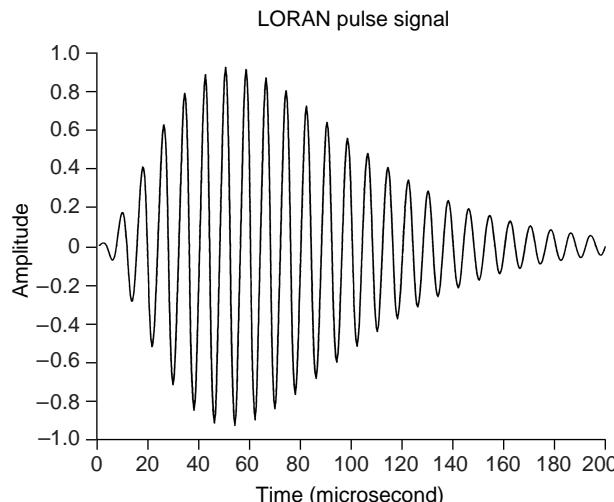


FIGURE 8 LORAN signal waveform.

Figure 8 depicts a typical LORAN current pulse. These pulses are periodically transmitted at an interval known as the group repetition interval (GRI), which interval identifies the chain. Each transmitting station in a chain has a specific signal format consisting of the number of pulses in a group, pulse spacing in a group, carrier phase code of each pulse, time of transmission, time between repetition of pulse groups, and delay of secondary station pulse groups relative to the master.

The LORAN-C signal propagates along the Earth's surface for relatively large distances. At the relatively low carrier frequency employed, the phase and group velocities can differ substantially owing to the complicated nature of the propagation velocity. The corresponding velocity variations result in temporal variations in the pulse time of arrival that have diurnal and seasonal components.

In addition, the local Earth spatial conductivity variations result in propagation time errors. Conductivity variations of 3 or 4 orders of magnitude occur for signal paths over seawater or over arid landmasses. It is beyond the scope of this article to further discuss the propagation errors or techniques for their partial compensation. Interested readers should consult Ref. (1) for a more complete discussion.

LORAN operation is based upon measurements of the time delay of the signal arrival from the master to the arrival time of the corresponding pulse from the secondary transmitters. The LORAN receiver determines the propagation time difference (TD) between these two signals. In the absence of any error source and within the validity of the planar approximation of the Earth's surface, we have

$$TD = \frac{1}{C} \left\{ \sqrt{\left(\frac{D}{2} + x \right)^2 + y^2} - \sqrt{\left(\frac{D}{2} - x \right)^2 + y^2} \right\},$$

where C = speed of propagation of LORAN signal
 D = separation of master and secondary

For any given TD, the above equation describes the locus of points (LOP) for constant TD. The curve $\pm y(x)$ satisfying this equation is a hyperbola. That is, a single measurement of the TD for a master and any one of its secondary transmitters (i.e., X , Y , and Z) locates the aircraft along a hyperbola.

A second measurement of the time delay between the master and another secondary TD₂ results in the location of the aircraft along a second hyperbola. In principle, the aircraft x , y coordinates can be found by solving a pair of equations in these two unknowns (i.e., the intersection of the hyperbolas). In practice, a LORAN receiver performs optimal statistical analysis on multiple measurements of the TD from multiple master-secondary stations.

The time of arrival of a pulse from a given transmitter within a chain is taken as the positive zero crossing of the pulse on its third cycle (i.e., approximately 30 μ s after pulse initiation). A modern aircraft LORAN receiver is under control of a microprocessor and associated programs. Under program control the receiver operates in four modes: (1) initialization, (2) acquisition, (3) pulse group time reference (PGTR) identification, and (4) tracking. During initialization, the receiver sets parameters that will generally optimize its performance. This process also may include adjustment/tuning of dynamically programmable interference rejection filters. Acquisition is the process of searching for and locating signals that have been identified during initialization. Essentially, the receiver "locks on to" a set of pulses such that the TDs between master and secondary are obtained with the greatest possible accuracy available at the operating conditions. The PGTR mode essentially minimizes various propagation errors. Tracking is the process of maintaining a fixed synchronization of the receiver with the PGTR for each signal being tracked. Additional details of the operation of LORAN can be found in Refs. (2–4).

VI. AUTOMATIC DIRECTION FINDING

One of the earliest radio navigation methods is radio direction finding. This method is based upon the directional sensitivity of a loop antenna. Consider an antenna consisting of a loop of conductor in a vertical plane. For a given radio transmitter at a specific distance the antenna terminal voltage is proportional to $\cos \theta$ where θ is the angle measured in a horizontal plane from the plane of the coil to a line through the coil center and transmitter location.

Direction finding was accomplished in an early aircraft (as well as in ships) by manually rotating the antenna until

a null terminal voltage was obtained. Of course, there is an ambiguity of 180° in this angular reading since the transmitter can be anywhere along the line orthogonal to the coil when the terminal voltage null is obtained. In early receivers the transmitter location was obtained by triangulation with a second receiver.

In a modern Automatic Direction Finding (ADF) receiver the single rotatable loop can be replaced by a pair of coils in orthogonal vertical planes. An ADF receiver system measures and displays the bearing of a station relative to the aircraft axes without operator intervention (other than tuning the receiver to the desired station transmitting frequency). The predecessor to ADF required an operator to perform several manual steps to achieve this result. The transmitter location can be found by a comparison of the relative amplitudes of the terminal voltages from the two coils. The 180° ambiguity is resolved by the use of a separate sense coil that provides a carrier phase reference to uniquely locate the transmitter.

VII. LONG-RANGE NAVIGATION

For navigation using Earth-based radio systems, ordinary charts that represent navigational coordinates on a flat plane approximation are adequate. Such navigation takes place over sufficiently short distances that this approximation is valid.

Long-range navigation using satellite-based navigation systems or inertial platforms require a more accurate representation of the three-dimensional (3-D) geometry of the Earth.

The Earth is essentially an ellipsoid of revolution about its spin axis. The semimajor and semiminor axes (a, b) of the Earth are

$$a = 6378.137 \text{ km} = 3443.918 \text{ nmiles}$$

$$b = 6356.752 \text{ km} = 3432.371 \text{ nmiles}$$

The ellipticity f and eccentricity e are

$$f = \frac{a - b}{a} = 0.003353$$

and

$$e = \frac{\sqrt{a^2 - b^2}}{a} = 0.08182.$$

In navigational problems, it is common to define the radii of curvature R_m and R_p , where R_m is the meridian radius of curvature of the best fitting circle to a meridian section of the ellipsoid (i.e., along a plane through the spin axis) and R_p is the prime radius of curvature which is the radius of the best fitting circle to a vertical east-west section of the ellipsoid. These radii are given by

$$R_m = \frac{a(1 - e^2)}{(1 - e^2 \sin^2 F_T)^{3/2}} \cong a \left[1 + e^2 \left(\frac{3}{2} \sin^2 F_T - 1 \right) \right]$$

and

$$R_p = \frac{a}{(1 - e^2 \sin^2 F_T)^{1/2}} \cong a \left[1 + \frac{e^2}{2} \sin^2 F_T \right],$$

where F_T = the geodetic latitude at a given location on the Earth's surface. This latitude is the angle between the normal to the meridian ellipse and the equatorial plane and is the latitude used on maps. The longitude λ is the angle measured from the prime meridian (through Greenwich, England) to the meridian at any given point.

The two radii of curvature (R_m and R_p) relate horizontal components of velocity to the angular coordinates:

$$\dot{F}_T = \frac{V_N}{R_m + h}$$

$$\lambda \cos F_T = \frac{V_E}{R_p + h},$$

where V_N is the component of horizontal velocity north, V_E is the east component of velocity, and h is altitude.

VIII. NAVIGATIONAL COORDINATE SYSTEMS

There are numerous coordinate systems for atmospheric flight navigation. The Earth centered-Earth fixed (ECEF) system is a set of Cartesian coordinates y_i , ($i = 1, 2, 3$) in which y_3 lies along the spin axis and y_1 lies in the prime meridian. These coordinates are typically used in satellite-based navigation systems.

Another major coordinate system is the Earth centered inertial (ECI) system which is similar to the ECEF except that y_1 is nonrotating relative to the fixed stars. This system is normally used in implementing an inertial navigation system. In this system, Newtonian mechanics are valid and angular coordinates of stars are tabulated in ECI coordinates.

Geodetic spherical coordinates Z_i ($i = 1, 2, 3$) are spherical coordinates in which Z_1 is longitude λ , Z_2 is geodetic latitude, and Z_3 is directed radially outward from the best fitting meridional circle.

IX. SATELLITE-BASED NAVIGATION SYSTEM

Radio-based navigation is also based upon the use of signals transmitted from satellites. There are two such systems in service: (1) the U.S. Department of Defense (DOD) NAVSTAR global positioning system (GPS) and

(2) the Russian Federation global orbiting navigation satellite system (GLONASS). The GPS system consists of 24 satellites arranged in groups of 4 in each of 6 orbital planes inclined at 55° spaced 60° apart in longitude and at a nominal altitude of 11,000 nmiles above the local surface (i.e., semimajor axis ≈ 26,600 km).

Each satellite carries a precise (atomic) clock and repetitively transmits its position and time. The user equipment consists of a receiver along with its own precise clock. By measuring the time difference δt from transmission of the signal to its reception, the receiver obtains a measurement of the transit time from satellite to receiver, which yields range R .

$$R = C\delta t = \sqrt{(x - x_s)^2 + (y - y_s)^2 + (z - z_s)^2}$$

where C is the speed of propagation of the satellite-transmitted signal and where x_s , y_s , z_s are the coordinates of the satellite. If the receiver and satellite clocks were perfectly synchronized then, in principle, measurement of δt would yield the range from the receiver to the (known) satellite position. A set of three measurements to three satellites could ideally yield the solution for the user position (x , y , z) from these measurements. However, it is, in practice, impossible to exactly synchronize these two clocks. The actual measured time difference between satellite (i) clock and receiver clock time yields an estimate of R (denoted R_i) called pseudo-range. Because of the receiver clock uncertainty, at least four measurements are required to estimate position and receiver clock error. The pseudo-range model is given by

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

where B is a bias resulting from the receiver clock error Δt_c :

$$B = C\Delta t_c,$$

that is, Δt_c is the error between true GPS time as carried by the satellite and the receiver clock time and C is propagation speed.

The user position can be determined (i.e., the navigation problem can be solved) by measurement of the pseudo-range to at least four satellites by triangulation or trilateration. However, the accuracy of this position solution is influenced by many factors, including the geometry of the satellites in relation to the user and various error sources. These errors include intentional degradation by DOD (significantly reduced in the year 2000) meant to reduce the accuracy to unfriendly users, propagation errors, clock random errors, orbital perturbations, and satellite ephemeris errors. Consequently, many more than four measurements are made to reduce errors. It is beyond the scope of this article to discuss the details of these error sources, and interested readers should consult Refs. (5–7).

In addition to estimations of position and clock bias, GPS receivers also estimate user velocity and clock drift rate. These additional estimates are required on an aircraft application since, in this case, the GPS receiver is presumed to be moving and to have changed its position considerably during the time required to obtain the pseudo-range measurements.

The procedure for estimating aircraft position and clock bias is to assume an initial position for the receiver and clock bias (x_o , y_o , z_o , B_o) and to find the aircraft estimated position at time t_k ($t_k = kT$, $k = 1, 2, \dots$):

$$\begin{bmatrix} x_o \\ y_k \\ z_z \\ B_k \end{bmatrix} = \begin{bmatrix} x_o \\ y_o \\ z_o \\ B_o \end{bmatrix} + \begin{bmatrix} \delta x \\ \delta y \\ \delta z \\ \delta B \end{bmatrix} + (k-1)T \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \\ \dot{B} \end{bmatrix},$$

where δx , δy , δz , δB are errors in the initial estimates, $(\dot{x}, \dot{y}, \dot{z})$ are the estimated user velocity vector components, and \dot{B} is the user clock drift rate.

The GPS navigation problem can now be formulated as a state estimation problem in which the state vector, X , is

$$X = [\delta x, \delta y, \delta z, \delta B, \dot{x}, \dot{y}, \dot{z}, \dot{B}].$$

The standard method for solving this problem is to linearize the pseudo-range equation:

$$R_i = R_{io} + \alpha_{i1}\delta x + \alpha_{i2}\delta y + \alpha_{i3}\delta z + \delta B,$$

where

$$\begin{aligned} R_{io} &= \text{initial pseudo-range estimate} \\ &= \sqrt{(x_o - x_i)^2 + (y_o - y_i)^2 + (z_o - z_i)^2} + B_o \end{aligned}$$

and where

$$\begin{aligned} \alpha_{i1} &= \left. \frac{\partial R_i}{\partial x} \right|_{R_o} = \frac{x - x_i}{R_{io} - B_o} \\ \alpha_{i2} &= \left. \frac{\partial R_i}{\partial y} \right|_{R_o} = \frac{y - y_i}{R_{io} - B_o} \\ \alpha_{i3} &= \left. \frac{\partial R_i}{\partial z} \right|_{R_o} = \frac{z - z_i}{R_{io} - B_o} \end{aligned}$$

The parameters α_{i1} , α_{i2} , α_{i3} are the direction cosines of the angles between the line of sight from the user to satellite i and the coordinate axes. The linearized pseudo-range equation can be written in terms of δR_i where

$$\delta R_i = R_i - R_{io} = \alpha_{i1}\delta x + \alpha_{i2}\delta y + \alpha_{i3}\delta z + \delta B.$$

[Figure 9](#) is a simplified illustration of the geometry for an aircraft moving at a constant velocity beginning at true position, x_T , y_T .

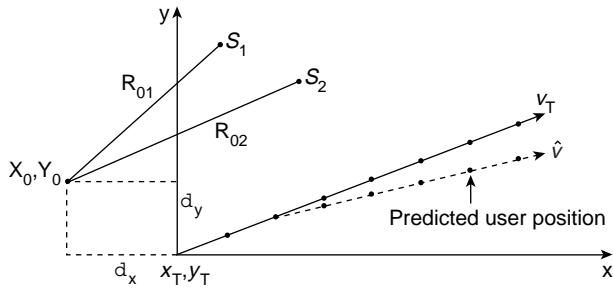


FIGURE 9 GPS navigation geometry (simplified).

The measurement model for the aircraft traveling at a constant speed from an initial estimated position (x_o, y_o, z_o) for four satellites is given by

$$\delta R = HX + e,$$

where

$$\delta R = \begin{bmatrix} \delta R_{11} & \delta R_{21} & \delta R_{31} & \delta R_{41} \\ & \vdots & & \\ \delta R_{n1} & \delta R_{n2} & \delta R_{n3} & \delta R_{n4} \end{bmatrix}^T.$$

In this expression n is the number of measurements made to each of the four satellites, and the matrix H is given by where

$$H = \begin{bmatrix} H_1 \\ H_2 \\ \dots \\ H_{n-1} \end{bmatrix},$$

where

$$H_k = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & 1 & kT\alpha_{11} & kT\alpha_{12} & kT\alpha_{13} & kT \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & 1 & kT\alpha_{21} & kT\alpha_{22} & kT\alpha_{23} & kT \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & 1 & kT\alpha_{31} & kT\alpha_{32} & kT\alpha_{33} & kT \\ \alpha_{41} & \alpha_{42} & \alpha_{43} & 1 & kT\alpha_{41} & kT\alpha_{42} & kT\alpha_{43} & kT \end{bmatrix}$$

$$k = 0, 1, \dots, n-1.$$

The state vector X is given by

$$X = [\delta x \ \delta y \ \delta z \ \delta B \ \dot{x} \ \dot{y} \ \dot{z} \ \dot{B}]^T,$$

and e is the $4n \times 1$ dimensional error vector.

The solution to this problem for finding X can be obtained in a “batch” mode, in which data is collected for n samples and found as the ordinary “least-squares” (OLS) solution, or the solution can be found recursively. The OLS solution is given by

$$X = (H^T H)^{-1} H^T \delta R.$$

Once a solution has been reached for X , the position of the aircraft for all time x_k, y_k, z_k is given by

$$\begin{bmatrix} x_k \\ y_k \\ z_k \\ B_k \end{bmatrix} = \begin{bmatrix} x_o \\ y_o \\ z_o \\ B_o \end{bmatrix} + \begin{bmatrix} \delta_x \\ \delta_y \\ \delta_z \\ \delta_B \end{bmatrix} + (k-1)T \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \\ \dot{B} \end{bmatrix}$$

for aircraft moving at a constant velocity vector $v = [\dot{x} \dot{y} \dot{z}]^T$. For maneuvering aircraft the solution involves time-varying estimates of V .

The recursive solution to the GPS navigation problem is a Kalman filter which continuously estimates X . The k^{th} estimate of X (i.e., X_k) is based upon K previous measurements and estimates and is of the form

$$X_k = X_{k-1} + K_k(\delta R_k - H_k X_{k-1}),$$

where K_k is the Kalman filter gain and δR_k is defined above. The Kalman filter gain is computed in a multistep process that is derived from known statistics of process and measurement noise, the basic dynamics of the aircraft motion, and the relationship between the state and the measurements in the zero noise ideal case. The details of these relations can be found in Refs. (8–10).

X. THE GPS SYSTEM STRUCTURE

The structure of the GPS navigation system consists of three major segments: (1) the space segment (the satellites); (2) the control segment; and (3) the user receiver systems. The satellite must be capable of transmitting its position and the correct GPS time continuously. A major function of the control segment is to periodically upload to each satellite data from which this position can be computed. Periodic updates to this ephemeris data are required owing to orbital perturbations and changes due to lunar–solar perturbations, air drag, asphericity of Earth’s gravitational potential, and magnetic static-electric forces in orbit.

The control segment configuration is depicted in Fig. 10. The monitor stations receive the GPS signals (same as the user). These signals can be used to evaluate ephemeris errors and satellite clock errors. These stations are located at Colorado Springs, Kwajalein, Diego Garcia, Ascension Island, and Hawaii. These stations measure pseudo-range values from the satellites as they come into view. These measurements are used to determine ephemeris and clock errors. In addition, these stations monitor local meteorological data that is useful for correcting for tropospheric delays. The data and corrections obtained by these monitor stations are sent to the master control.

The master control uploads navigation messages to the satellites via the stations at Ascension, Diego Garcia, and Kwajalein. The satellites are continuously controlled via

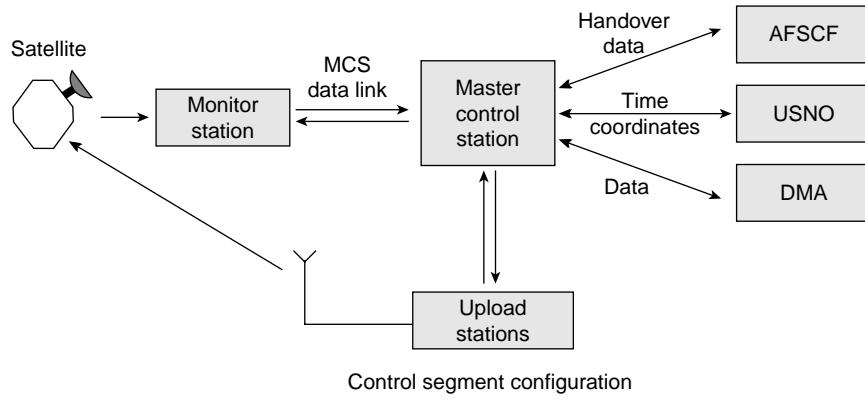


FIGURE 10 Control segment configuration.

the master control to avoid cumulative errors that would occur in the absence of this control function.

There are numerous error sources in GPS navigation solutions, including satellite ephemeris errors, propagation errors and uncertainties, and clock errors. These errors are exacerbated by poor geometry, which increases the uncertainty in position. Such uncertainty is represented quantitatively by a parameter known as geometric dilution of position (GDOP).

The ephemeris errors result from imperfect prediction of satellite position. Propagation errors and uncertainties result from ionospheric and tropospheric refraction variation. The ionospheric refraction is determined largely by free-electron density and carrier frequency. The index of refraction, n , for propagation through the ionosphere is defined as

$$n = \frac{C}{v_\phi},$$

where v_ϕ is the phase velocity.

At any carrier frequency, f , the index of refraction is given by

$$n = \sqrt{1 - \left(\frac{f_c}{f}\right)^2},$$

where f_c is the plasma frequency

$$f_c = \frac{1}{2\pi} \sqrt{\frac{N_e e^2}{m \epsilon_0}} \cong 9\sqrt{N_e}$$

where N_e = electron density (number/m³)
 e = electron charge
 m = mass of electron
 ϵ_0 = permittivity of free space

On the other hand, tropospheric refraction is independent of carrier frequency, but is influenced by the partial

pressure of water vapor in an approximate relationship [Refs. (11 and 12)] as represented by index of refraction, n :

$$n \cong 1 + \frac{K_1}{T} \left(P + \frac{K_2 P_w}{T} \right)$$

where P = atmospheric pressure
 T = absolute temperature
 P_w = partial pressure of water.
 $K_{1,2}$ = constants

The path length change due to refraction ΔL is given by

$$\Delta L = \int_0^R (n - 1) ds,$$

where R = distance to the satellite
 \cong pseudo-range
 s = coordinate along the propagation path

This expression can be rewritten approximately in terms of altitude h and elevation angle ϕ_o to the satellite:

$$\Delta L = \int_{h_o}^H \frac{n - 1}{\sin \phi_o} dh,$$

where H is the satellite elevation above the Earth. If the atmosphere is assumed to be exponential, then

$$n - 1 \cong (n_o - 1)e^{-bh},$$

where (for a standard day)

$$n_o \cong 1.00032$$

$$b \cong 0.000145/\text{meter}.$$

Typical values are

$$\begin{aligned} \Delta L &\cong 2.2 \text{ m} & \text{for } \phi_o &= 90^\circ \\ &\cong 25 \text{ m} & \text{for } \phi_o &= 5^\circ, \end{aligned}$$

The influence of satellite geometry is given via the GDOP. The GDOP can be computed from the matrix of direction cosines to the satellite, H_i , where the i^{th} row of H is

$$H_i = [\alpha_{i1} \ \alpha_{i2} \ \alpha_{i3} - 1]$$

$$\text{GDOP} = \sqrt{\text{trace}\{[H_i^T H_i]^{-1}\}}.$$

Then the rms error is given by

$$\sigma = \text{GDOP } \sigma_o,$$

where σ_o is the minimum position error that results for optimal satellite geometry and is due to the error sources listed above.

XI. WAAS

The accuracy of GPS navigation can be improved by a scheme that, at least partially, corrects common errors (bias errors) due to causes outside of the receiver. These error sources include (1) errors that are intentionally introduced by the DOD, (2) ionospheric delays, (3) tropospheric delays, (4) ephemeris errors, and (5) satellite clock errors.

XII. GPS ACCURACY AUGMENTATION

The accuracy improvement is achieved via a method known as differential GPS or (DGPS). The concept for DGPS involves a reference GPS receiver located at a precisely surveyed site that can calculate errors in pseudo-range and pseudo-range rate. These errors can then be transmitted via a suitable carrier frequency to all participating users within a specific geographic area.

DGPS methods are commonly known as augmentation schemes and are generally divided into two types depending upon the size of the geographic area covered by the system. These DGPS augmentation systems are (1) wide area augmentation systems (WAAS) and (2) local area augmentation systems (LAAS).

The WAAS system consists of a master station (WMS), several reference (monitoring) stations (WRS), and a number of geostationary communication satellites (GEO). The reference stations all continuously track the GPS and GEO satellites and relay tracking information (via land links) to the master station. Error corrections are determined by calculation and are relayed to users via the GEO. In this system, the corrections are transmitted on one of the GPS carrier frequencies.

LAAS augmentation involves a multitude of receivers, each serving a local area. Errors detected by these receivers

are then transmitted to all users in the vicinity of that receiver. LAAS, if implemented, should yield much higher accuracy than WAAS, although it will require a larger infrastructure investment.

XIII. GPS INTEGRITY AND AVAILABILITY

It has been proposed that GPS be used as the sole navigation means for the U.S. national airspace system. In order for this to be achieved, there are several major issues to be considered, including (1) accuracy, (2) integrity, (3) availability, and (4) continuity. The accuracy must be sufficient to safely achieve the navigational function, particularly with respect to instrument approaches in inclement weather. System integrity is the capability of the system to detect anomalous signals that could cause errors in excess of that allowed for safe use of the system. Availability means that the signal is present at a level necessary to provide the navigational function for the entire duration of the relevant portion of the flight. Continuity means that the navigational signal is present at the required level without interruption during the flight.

At the present time, GPS by itself does not satisfactorily resolve these issues. However, the FAA has defined three types of services related to GPS: (1) a multisensor system, (2) a supplemental system, and (3) a required navigation performance (RNP) system. In a multisensor navigation system, GPS with any available augmentation (e.g., DGPS) can be used for navigation only after it has been compared for integrity with another approved navigation system (e.g., VOR) on board the aircraft. In a supplemental system, the GPS (with augmentation) may be used without comparison to another approved navigation system, but another approved navigation system must be available on the aircraft and usable in case GPS is not available.

An RNP system is one that satisfies all requirements for the navigation phase without the need for any other navigation equipment on board. An RNP system may include one or more navigation systems, e.g., GPS with inertial navigation (see next section).

XIV. RECEIVER AUTONOMOUS INTEGRITY MONITORING (RAIM)

One of the important requirements for the GPS system to provide independent (stand-alone) navigation is the need for the system to continuously monitor the signal integrity. The method for achieving this is known as receiver autonomous integrity monitoring (RAIM). The goal of RAIM is to detect and isolate an out-of-tolerance signal that causes a navigational error in excess of that allowed

for the phase of flight in progress. It is assumed that the navigation solution has been obtained and that five or more satellites are in view.

The RAIM algorithm is based upon error residuals from the least-squares (or Kalman filter) solution. These error residuals have been denoted X above and are used to calculate the sum of squared errors (SSE).

$$\text{SSE} = X^T W X,$$

where W is a weighting matrix that is normally chosen in the design of the receiver (note: W could be selected to be the identity matrix). The SSE is a scalar error that generally is small for reliable signals and relatively large in the event of an anomalous signal condition. The GPS system has sufficient integrity for use if

$$\text{SSE} < \text{threshold value}$$

and is not suitable for use if

$$\text{SSE} > \text{threshold value}$$

where the threshold value depends upon the phase of flight in progress. That is, the threshold for approach and landing phases is smaller than for enroute navigation phases.

XV. INERTIAL NAVIGATION

Inertial navigation is a method of calculating the position and velocity of an aircraft based upon measurements of acceleration. Such a system can operate without external measurements from Earth or satellite radio signals, but it does require initialization along and around all axes. Velocity and position components are found by integrating each corresponding acceleration:

$$V_i = \int_0^t a_i(\tau) d\tau + V_i(0) \quad i = 1, 2, 3$$

$$x_i = \int_0^t V_i(\tau) d\tau + x_i(0),$$

where a_i = acceleration component along the i^{th} coordinate

V_i = velocity component along the i^{th} coordinate

x_i = position component along the i^{th} coordinate

Similarly, angular velocity and position can be found by integrating angular acceleration:

$$\begin{aligned} \omega_i &= \int_0^t \dot{\omega}_i(\tau) d\tau + \omega_i(0) \\ \phi_i &= \int_0^t \omega_i(\tau) d\tau + \phi_i(0) \quad i = 1, 2, 3 \end{aligned}$$

where $\dot{\omega}_i$ = angular acceleration about axis i
 ω_i = angular velocity about axis i
 ϕ_i = position about axis i

The implementation of inertial navigation is dependent upon the choice of navigational coordinates, as well as the aircraft altitude in relation to these coordinates, and the coordinates of the accelerometers. The navigational coordinates most commonly used for inertial navigation are the geodetic wander azimuth coordinates (GWA). This Cartesian coordinate system has z_3 vertically up and z_2 at an angle α relative to true north with $z_1 z_2$ forming a tangent plane to the Earth reference ellipsoid.

Inertial navigation systems operate by having a stabilized platform that maintains a fixed orientation in inertial space isolated from angular motion of the aircraft. The platform is either maintained physically via gimballed gyroscopes or analytically via computations made from gyroscopes (known as "strapdown" gyroscopes) that are physically attached to the aircraft structure.

For a gimballed gyroscope platform, the accelerometers have axes of known orientation. For strapdown systems, an equivalent platform orientation is computed from calculated coordinate positions.

Transformation of coordinates is via a matrix C of direction cosines. For example, the transformation of any vector V from ECEF coordinates to GWA coordinates is given by

$$\begin{bmatrix} V_{y1} \\ V_{y2} \\ V_{y3} \end{bmatrix} = C_{zy} \begin{bmatrix} V_{z1} \\ V_{z2} \\ V_{z3} \end{bmatrix} \quad z_i, y_i = 1, 2, 3,$$

where $C_{11} = \cos \alpha \sin \ell - \sin \alpha \sin F_T \cos \ell$
 $C_{12} = \cos \alpha \cos \ell - \sin \alpha \sin F_T \sin \ell$
 $C_{13} = \cos F_T \sin \alpha$
 $C_{21} = \sin \alpha \sin \ell - \sin F_T \cos \alpha \sin \ell$
 $C_{22} = \sin \alpha \cos \ell - \cos \alpha \sin F_T \sin \ell$
 $C_{23} = \cos F_T \cos \alpha$
 $C_{31} = \cos F_T \cos \ell$
 $C_{32} = \cos F_T \sin \ell$
 $C_{33} = \sin F_T$

where F_T = geodetic latitude
 ℓ = longitude
 α = wander azimuth angle

Conversion to ECI coordinates is straightforward, based upon knowledge of universal coordinated time (i.e., time at Greenwich meridian).

In a strapdown inertial navigation system, the vehicle attitude (body axes) is computed relative to navigation coordinates via direction cosine matrix $C_B^N(B, \text{body}; N, \text{navigation})$:

$$C_B^N = \begin{bmatrix} S_\psi C_\theta & S_\psi S_\theta S_\phi & S_\psi S_\theta C_\phi - C_\psi S_\phi \\ C_\psi C_\theta & C_\psi S_\theta S_\phi - S_\psi C_\phi & C_\psi S_\theta C_\phi + S_\psi S_\phi \\ S_\theta & -C_\theta S_\phi & -C_\theta C_\phi \end{bmatrix},$$

where

$$\left. \begin{aligned} C_i &= \cos(i) \\ S_i &= \sin(i) \\ i &= \psi = \text{yaw} \\ \theta &= \text{pitch} \\ \phi &= \text{roll} \end{aligned} \right\} \text{Euler angles}$$

This matrix is dynamic due to aircraft motion:

$$\dot{C}_B^N = C_B^N \omega_{BN}^N,$$

where ω_{BN}^N is the instantaneous angular rate vector of the body frame relative to the navigation frame in body coordinates.

Unfortunately, in practice, for transport aircraft this method of computing instantaneous direction cosines in strapdown systems suffers from errors called coning errors. An alternative formulation for strapdown is based upon quaternions. A quaternion is a four-element quantity having a scalar part λ and a vector part $\bar{\rho}$ and is denoted Q :

$$Q = [\lambda, \bar{\rho}].$$

For a rotation quaternion, Q becomes

$$Q = \left[\cos\left(\frac{\phi}{2}\right), \bar{u}_\phi \sin\left(\frac{\phi}{2}\right) \right],$$

where \bar{u}_ϕ is a unit vector along the inclined axis of rotation and ϕ is the rotation about that axis.

The quaternion has unity L_2 norm

$$\|Q\|_2 = \lambda^2 + \bar{\rho} \cdot \bar{\rho} = 1$$

and has inverse given by

$$\begin{aligned} Q^{-1} &= [\lambda, \bar{\rho}]^{-1} \\ &= [\lambda, -\rho]. \end{aligned}$$

The differential equation for a quaternion is

$$\frac{dQ}{dt} = \frac{1}{2} Q [0, \bar{\omega}],$$

where $\bar{\omega}$ is the instantaneous angular rate vector. The direction cosine matrix from body to navigation coordinates (properly normalized) is

$$C_B^N =$$

$$\begin{bmatrix} \lambda^2 + \rho_x^2 + \rho_y^2 + \rho_z^2 & 2(\rho_x \rho_y - \lambda \rho_z) & 2(\rho_x \rho_z + \lambda \rho_y) \\ 2(\rho_x \rho_z + \lambda \rho_y) & \lambda^2 - \rho_x^2 - \rho_y^2 - \rho_z^2 & 2(\rho_y \rho_z - \lambda \rho_x) \\ 2(\rho_x \rho_z - \lambda \rho_y) & (2\rho_y \rho_z + \lambda \rho_x) & \lambda^2 - \rho_x^2 - \rho_y^2 - \rho_z^2 \end{bmatrix}.$$

The mechanization of navigation in the horizontal (navigation coordinates) requires integration of the following:

$$\frac{dV_n}{dt} = \underline{u} - (\underline{\omega} + \underline{\Omega}) \times \underline{V} + \underline{g},$$

where \underline{V} = velocity vector

\underline{u} = acceleration vector (in navigation coordinates)

$\underline{\omega}$ = inertial angular velocity

$\underline{\Omega}$ = angular velocity of the Earth in inertial space

\underline{g} = local gravity vector

The middle term (vector product) in the above expression corresponds to Coriolis acceleration. The solution to the above equations yields the components of the ground velocity vector. The mechanization involves computation of the components of each of the terms in the above equation in navigation coordinates (i.e., in x_N , y_N , z_N), where the associated components are found using the appropriate direction cosine matrices. Figure 11 is a block diagram of a representative inertial navigation system (INS).

In this figure, we have only depicted the x_N coordinate subsystem. A similar subsystem applies for the y_N coordinate. The commonest choices of coordinates for inertial navigation are the wander azimuth coordinates.

There are numerous error sources that influence the accuracy of a stand-alone inertial navigation system, including gyro drift, gyro scale factor, accelerometer bias, accelerometer scale factor, sensor assembly alignment, computational errors, and initial condition. Unfortunately, if uncorrected, these errors can increase monotonically with time, possibly becoming significant for a long trip. However, a typical inertial navigation system with no error corrections can achieve an error of less than 1 nmiles in 1000 nmiles.

A trend for future inertial navigation systems is to integrate them into a multisensor environment. The addition of GPS for error correction has the potential to provide highly accurate long-range navigational accuracy and, at the same time, the security of a system that can function in poor or degraded GPS accuracy.

Owing to space limitations, this discussion of inertial navigation has been limited in scope. This is a relatively complex subject with a technology that has evolved over

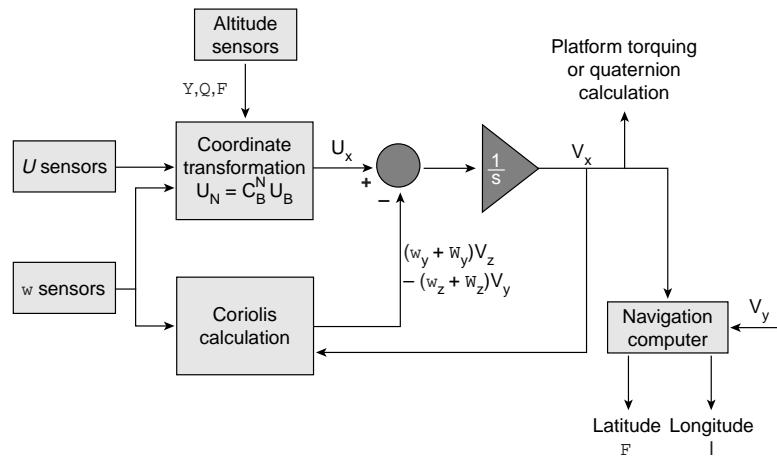


FIGURE 11 Representative inertial navigation system block diagram.

decades and continues to evolve. Interested readers should consult Refs. (15 and 16).

XVI. INSTRUMENT LANDING SYSTEM (ILS)

The phase of flight that requires the greatest navigational accuracy is the landing phase in poor weather conditions requiring IFR flight operations. Whenever the ceiling exceeds 1000 ft and the visibility 3 miles, landings can be conducted under VFR rules. If either of these conditions are not met, the landing must be performed under IFR. Instrument landings are either nonprecision, in which radio navigation aids provide lateral guidance (i.e., right or left of runway extended centerline), or precision, in which radio navigation aids also provide vertical guidance. Precision approaches fall under categories I, II, and III with associated minimum ceiling and visibility requirements. Category III ILS is further divided into IIIa, IIIb, and IIIc. Each of these categories has associated minimum visibility requirements that get progressively more stringent moving from category I through category IIIc. This latter category is associated with zero visibility and zero ceilings.

The standard minimums for a category I ILS are 200 ft ceilings and 1/2 mile visibility. For category II the minimums are 100 ft and runway visual range (RVR) of 1200 ft. However, special equipment and aircrew certification are required for a category II approach. Similar certification and special equipment is required for category III with zero ceiling and RVR of 700 ft for category IIIa and 600 ft for category IIIb. The category IIIc minimums are zero ceiling and RVR, but only a few airports and a small number of aircraft meet the requirements for category IIIc anywhere in the world.

Aircraft operating under IFR conditions transition from cruising flight to the final approach using procedures published for the airport. The final approach coarse is an extended runway centerline, and the last segment of the transition normally involves a heading change of about 20° to align the aircraft ground track with the final approach course.

Figure 12 illustrates the landing segment environment for a precision approach ILS.

The final approach course includes a line segment from the final approach fix (FAF) to the runway threshold. This line segment, which is known as the glide path, makes an angle with the local horizon known as the glide path angle. This angle is chosen for safe obstacle clearance with the standard value of 3°. The FAF is a point along this line segment that can be 4–7 miles from the runway threshold.

Perhaps the most common method of transitioning to the FAF is via the λa heading vector assigned by approach control. The aircraft will be given this assigned heading and an assigned altitude via a barometric instrument or altimeter, until the flight path intercepts the final approach course. The intercept altitude depends upon terrain and obstacle clearance, but in flat terrain it is typically more than about 1500 ft. In mountainous terrain, the aircraft may

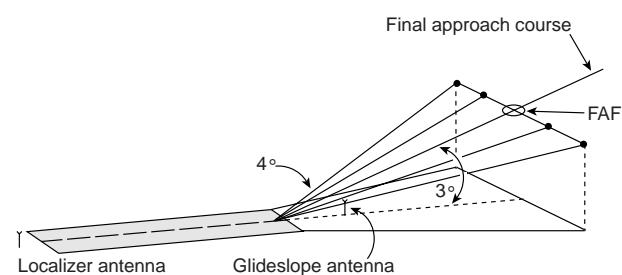


FIGURE 12 Illustration of geometry of an ILS.

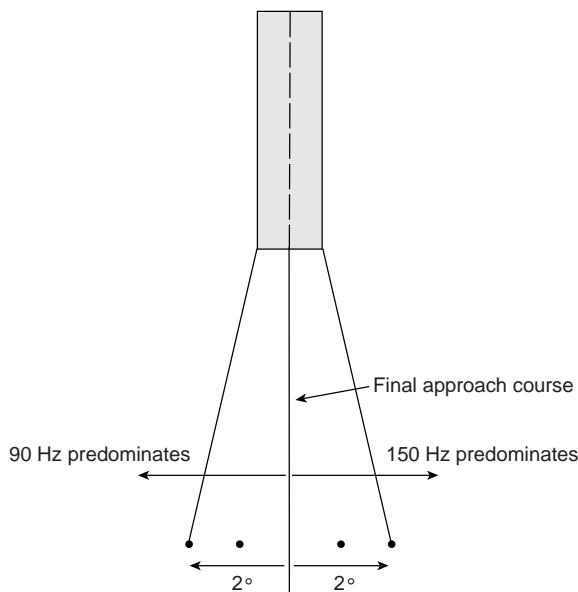


FIGURE 13 Illustration of localizer geometry.

intersect the final approach course well outside the FAF and several thousand feet above the runway threshold.

Once the aircraft has intercepted the final approach course, the aircraft is guided to the runway threshold by means of the ILS instrumentation working in conjunction with radio signals transmitted through antennas located near the runway. Lateral guidance is accomplished with a portion of this instrument known as the “localizer.” The localizer system consists of a transmitter and antenna system located at the upper end of the runway from the approach threshold as well as a receiver in the aircraft and cockpit display.

Vertical guidance is accomplished with a portion of the ILS known as the “glide slope,” which includes a transmitter and antenna located to the side of the runway about 1000 ft from the threshold as well as a receiver and cockpit display.

The localizer consists of a transmitter with carrier frequency in the range $108 \leq f_c \leq 112$ mHz. The carrier is

amplitude modulated at 90 and 150 Hz to form a carrier with sidebands (CSB) signal as well as a sideband-only (SBO) signal. These signals are combined with precisely controlled phase such that in the radiated pattern to the right of the final approach course the 150 Hz predominates and to the left the 90 Hz predominates, as illustrated in Fig. 13.

It is beyond the scope of this article to explain the details of the radiated pattern for the ILS, but interested readers should consult Ref. (13).

The localizer receiver block diagram is shown in Fig. 14. It is a conventional superheterodyne receiver that incorporates narrow band filters to separate the 90- and 150-Hz signals.

The filtered signals are detected and sent to a differential amplifier. The output of this amplifier drives the indicator (I), which is a galvanometer (or its functional equivalent) and could well be a part of a solid state digital display. The deflection of the pointer is proportional to the deviation of the aircraft from the final approach course. If the aircraft is to the right of course, the 150-Hz signal predominates and the galvanometer pointer is left of center indicating to the pilot to fly left and vice versa. If the aircraft deviates more than $\pm 2^\circ$ from the centerline, the indicator is saturated and the pilot is obliged to perform a missed approach, which essentially means climbing to a safe altitude and, under ATC control, instituting the approach again.

The glide slope is similar to the localizer in that a carrier is modulated with 90- and 150-Hz signals, as illustrated in Fig. 15. The carrier frequency is in the band 324–335 mHz.

The glide slope operates in much the same way as the localizer except that the tolerance bands from the glide slope are $\pm 0.7^\circ$.

The radiation pattern required for such angular selectivity is achieved by the antenna array and the precise phase relationships between the CSB and SBO signals. The international standards for ILS require that a full-scale deflection of the ILS cockpit display be provided to assist the flight crew in proper acquisition of the final approach course to $\pm 35^\circ$ of the final approach course. It should be

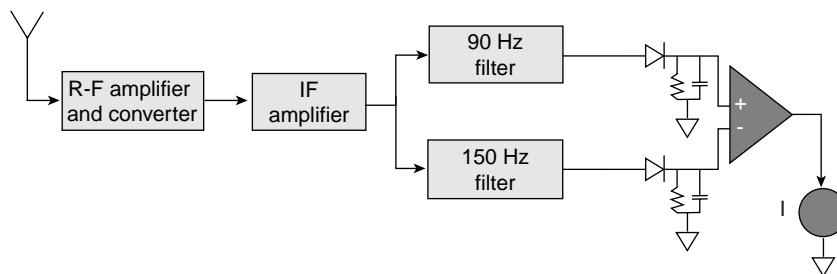


FIGURE 14 Block diagram for a localizer receiver.

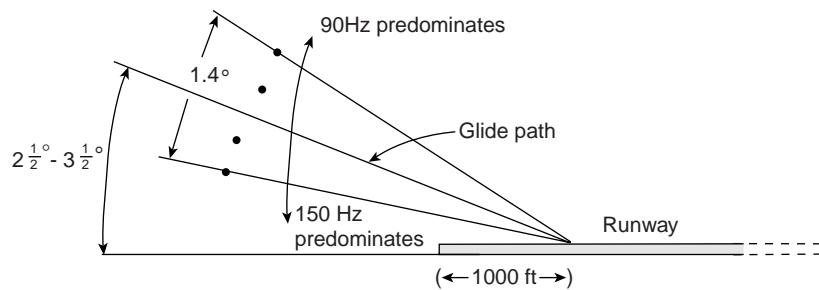


FIGURE 15 Illustration of glideslope geometry.

noted, however, that the display comes out of saturation only when the aircraft is within $\pm 2^\circ$ of this course.

There are a number of potential error sources that influence the accuracy and precision of the ILS, including multipath from reflections off large buildings. In order to achieve category II or III landings such errors must be reduced. Typically, such error reduction is achieved by reducing the CSB beamwidths 5° to 10° . The $\pm 35^\circ$ coverage requirement is achieved by providing broad beamwidth (CSB and SBO) signals transmitted on a separate but closely spaced carrier. This signal is known as the clearance signal and provides the ILS receiver with the acquisition capability to $\pm 35^\circ$. However, this signal is relatively weak such that when the aircraft is within the narrower beamwidth of the main CSB and SBO signals, these latter signals predominate and control the ILS display.

The antenna systems used for localizer and glide slope are partly responsible for achieving the radiation patterns required. The localizer uses a horizontally polarized log-periodic dipole antenna located 600–1000 ft beyond the stop end of the runway. Each element of this antenna is driven by a balanced transmission line fed from an impedance matching network yielding a unidirectional radiation pattern. For the glide slope, two or three antenna arrays are arranged on a vertical mast that is placed to the side of the runway about 1000 ft down the runway from the threshold. Details of the ILS antenna systems are given in Ref. (14).

XVII. ADDITIONAL PRECISION APPROACH SYSTEMS

In addition to ILS, precision approach navigation is potentially available using a microwave landing system (MLS), precision DME, and augmented GPS. Of these, only the GPS is likely to see widespread civil aviation use. While nonprecision GPS approaches are presently in use and approach procedures have been published for several airports, these require equipment with special IFR certification for IFR use.

At the present time, considerable effort is being expended to develop and release for commercial use category I DGPS precision landing systems. Category I precision approaches can potentially be achieved using the WAAS augmentation system (see section on GPS navigation). Moreover, category II and III landing systems require a local differential augmentation. However, for any such system, the required accuracy of the GPS measurements must be about ± 8 m for category I, ± 4 m for category II, and ± 1.3 m for category III. Experimental category IIIc automatic landings have been demonstrated, but widespread commercial use will not occur until some time in the future.

Precision GPS approaches have certain potential advantages over ground-based radio landing systems in that a multisegment approach can be designed for terrain avoidance in mountainous regions. In principle, such an approach could have one or more curved arcs with connected straight segments involving heading changes at each intercept. However, in practice, aircraft dynamics, terrain features, and other factors limit the achievable approach paths somewhat.

At one point, it was suggested that GPS, augmented suitably, could provide a sole basis for all aircraft navigation. However, potentially serious signal integrity questions have arisen that have reduced the initial enthusiasm for GPS-only navigation. Included in the threats to signal integrity are both unintentional and intentional interference to GPS signals from various radiofrequency sources. Nevertheless, GPS will increasingly become the dominant navigation means (including landing phase) over the near future.

XVIII. AIR DATA SYSTEM

Measurement of critical flight variables such as airspeed and altitude have long been important in aircraft. In lower performance general aviation aircraft, such measurements are still today performed by stand-alone pneumatic-mechanical instruments that respond as required to static,

dynamic, or total pressure. In high-performance (and high cost) general aviation, transport, and military aircraft, these and other variables must be computed to relatively high accuracy and must be available in a computer-based instrument where these variables can be combined in known functional relationships to evaluate and optimize aircraft performance.

An air data system provides calculations of flight variables, including calibrated airspeed, true airspeed, equivalent airspeed, Mach number, free-stream static pressure and outside air temperature, air density, pressure altitude, density altitude, angle of attack, and side slip angle. The static pressure p_s is the absolute pressure of the still air at any point in the atmosphere. An approximate measurement of static pressure can be obtained by means of a port along the side of the fuselage (called the static port). Total pressure p_t is the pressure sensed in a tube that is open at the front, closed at the rear, and directed toward the free-stream air velocity vector.

The various airspeeds are derived by computation from measurements of total pressure, static pressure, and absolute air temperature T . Impact pressure q_c is defined as

$$q_c = p_t - p_s,$$

which for subsonic flight is given by

$$q_c = p_s \left[1 + \left(\frac{\gamma - 1}{\gamma} \right) \frac{\rho}{2p_s} V^2 \right]^{\frac{1}{1-\gamma}} - p_s,$$

where ρ = local air density (slug/ft³)
 V = true airspeed(ft/sec)
 γ = ratio of specific heats for air = 1.4

Air density can be obtained from local air static pressure and temperature

$$\rho = \frac{p_s}{gRT},$$

where R = 53.3 ft per degree Kelvin
 g = acceleration of gravity

The true airspeed can be obtained by solving the q_c equation for V . Mach number M is the ratio of V to the local speed of sound a :

$$M = \frac{V}{a},$$

where

$$a = \sqrt{\gamma g RT}.$$

Calibrated airspeed is the value that would be obtained from the q_c equation if measurements were made at sea level on a standard day at which $\rho = \rho_0 = 0.002378$ slug per cubic feet.

The above measurements of speed also require a measurement of local outside air temperature. A temperature sensor mounted on the outside of an aircraft measures a temperature that is higher than the still air temperature due to friction and compressibility. The measured temperature T_M is given by

$$T_M = T \left[1 + \frac{\gamma - 1}{2} \eta M^2 \right],$$

where η is the empirically determined constant for the sensor.

The air data system solves this equation for T from measurements of T_m and M .

Altitude measurements are derived from the measurement of static pressure and a standard model for the atmosphere in the following equations:

$$\frac{dp_s}{dh} = -g\rho$$

$$\bar{T}(h) = T_o - \alpha_a h$$

where h = true altitude
 α_a = lapse rate for the atmosphere
 $= 0.003566^\circ\text{F}/\text{ft}$

Integrating the above equation in conjunction with the air density equation yields the standard atmosphere model for pressure vs altitude:

$$p = p_0 \left(1 - \frac{\alpha h}{T_o} \right)^{\frac{1}{\alpha R}}.$$

Altitude can be computed using this relationship and measurements of static pressure and outside air temperature. Angle of attack and side slip angle are measured using rotary position sensors connected to movable vanes that are mounted on the surface of the aircraft as illustrated in Fig. 16. The rotary position sensor is essentially a potentiometer having a movable vane attached to the rotary shaft. The movable vane aligns with the air velocity vector. The angle between this vane and the aircraft longitudinal axis is the angle of attack. Consequently, the output voltage from the potentiometer is an essentially linear function of angle of attack.

The vane assembly is mounted flush with the aircraft surface via a mounting flange. The plane of symmetry of the flange is in the plane of symmetry of the aircraft for

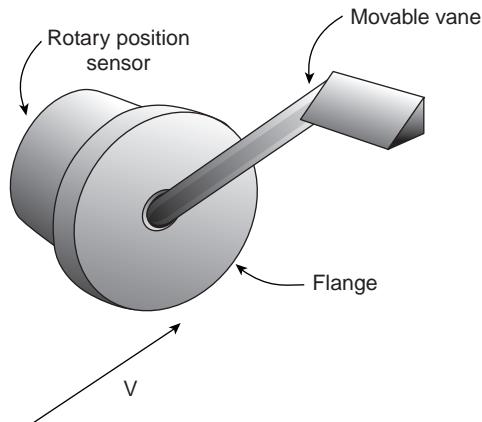


FIGURE 16

side slip measurements and is in a horizontal plane for angle of attack measurements.

A block diagram of an air data system is depicted in Fig. 17.

Sensors for measuring pressure are available in a variety of technologies that often incorporate a diaphragm that seals a closed chamber and is coupled to a displacement sensor. Another class of pressure sensor involves fabricating the diaphragm from doped silicon whose resistivity varies with stress due to a piezo-resistive property. Stress dependent resistance is readily converted to a measurement of pressure for a given diaphragm configuration via a bridge circuit or the like.

Temperature sensors often consist of a small coil of wire whose resistance varies with temperature. Alternatively, a semiconductor slab can also provide a temperature dependent resistance. Water ingestion and icing can cause significant errors that must be minimized by design.

The calculated values for the various air data variables are used in cockpit instrumentation via a suitable display such as an analog or digital display or a cathode ray tube or solid state equivalent.

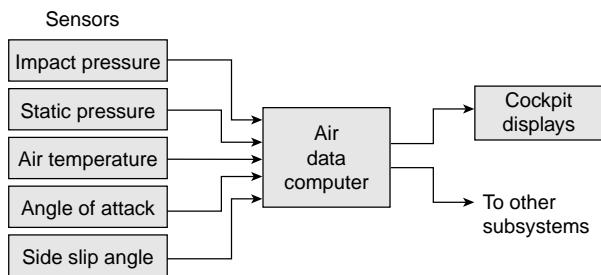


FIGURE 17 Air data system block diagram.

XIX. ATTITUDE AND HEADING REFERENCES

Safe flight in instrument meteorological conditions (IMC) requires instrumentation for measuring the aircraft altitude, i.e., pitch, roll, and yaw angles relative to the locally level plane. Roll and pitch are measured and displayed in the cockpit on a single instrument (called the attitude indicator), and yaw in the form of a heading angle is displayed by an instrument known as directional gyro or DG (from its earliest implementation).

In most general aviation and older transport and military aircraft, altitude and heading systems (AHRS) are via gimballed gyroscopes. In modern transport military and high-performance general aviation aircraft, the AHRS is an electronic instrument that uses gimballed gyroscopes with rotary position sensors to measure the angular displacement between the gyro axis and gimbal system which can be related directly to the aircraft body axes.

A functional block diagram of an electronic AHRS is depicted in Fig. 18. The sensors for pitch and roll are part of a gyroscope system having a vertical spin axis. The pitch and roll rotary position sensors are mounted along two orthogonal gimbals. The heading sensor measures the angular displacement between a horizontal gimbal and the axis of a gyroscope whose spin axis is horizontal.

The signal processing is implemented with a digital computer in modern AHRS systems that has the capability of system initialization and error compensation and generates signals to drive the display.

Clearly, initialization of these gyroscopes and the corresponding sensors is critical for accuracy of these instruments. Part of the initialization involves the orientation, during start-up, of the vertical gyroscope (VG) spin axes in the local vertical (corresponding to the gravity vector). Most commonly, this so-called gyro-erection process involves some form of pendulum and takes place with the aircraft at rest on the ground. When AHRS is combined with an inertial navigation system (either platform or strapdown type), the local vertical is available by appropriately correcting the net vehicle acceleration vector for transverse acceleration components.

Typically, leveling with a pendulum involves torqueing the gyroscope in a direction that causes the spin axis to precess toward true vertical as depicted in Fig. 19.

The dynamics of this system are given by

$$\dot{\varepsilon} + \frac{k}{H}\varepsilon = \dot{\theta} + \omega_D,$$

where H = gyro angular momentum
 k = sensitivity of torque to tilt

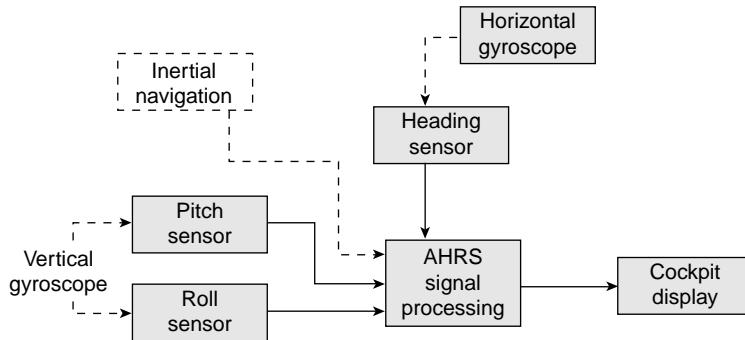


FIGURE 18 Typical AHRS block diagram.

$$\begin{aligned}
 \omega_D &= \text{drift rate for gyroscope} \\
 \dot{\theta} &= \text{angular rotation rate of the apparent} \\
 &\quad \text{vertical in inertial coordinates} \\
 \varepsilon &= \text{alignment error}
 \end{aligned}$$

For small angular displacements, the following approximation is valid:

$$\begin{aligned}
 \sigma &= \varepsilon + \delta \\
 &\approx \varepsilon + \frac{a_H}{g} \quad (\text{see Fig. 19})
 \end{aligned}$$

The system dynamics are given by

$$\dot{\sigma} + \frac{k}{H}\sigma = \frac{a_H}{g} - \dot{\theta} + \omega_D.$$

The response of this control system (which is first order) is stable, but it will not have asymptotically zero error (i.e., $\varepsilon \rightarrow 0$). The system for aligning the gyro axis with true vertical includes a control system that applies a torque to the gyro gimbal causing the error to go to zero. In a simple system the torque is proportional to error ε . However, the response of this control system can be tailored to achieve zero asymptotic error if a term proportional to $\int \varepsilon dt$ is included.

Horizontal acceleration results from aircraft acceleration (relative to the Earth) as well as Coriolis acceleration. For an aircraft in steady, level flight at ground speed, V_g , and latitude, ϕ , a_H is given by

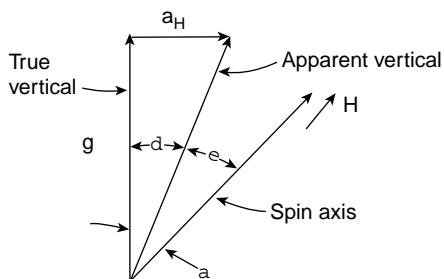


FIGURE 19 Illustration of geometry for alignment of the VG axis.

$$a_H = 2\Omega V_g \sin \phi,$$

where Ω is Earth's angular velocity.

A horizontal reference is required for initialization as well as error compensation for the heading indicator. For those aircraft equipped with an inertial navigation system such error compensation is available as one of the system outputs. For any other aircraft the best available heading correction and initialization is via Earth's magnetic field. The flux density vector for Earth's magnetic field, \bar{B}_E , is that of a dipole oriented at 11° to the spin axis. The axis of this dipole field passes essentially through the center of the Earth. Although it moves around somewhat, the magnetic pole lies roughly in the meridian that is about 84° to 85° west longitude. There are numerous analytical models of the field distribution of \bar{B}_E that are periodically updated. Essentially, the field strength is about 0.3 G near the equator to about 0.6 G near the poles. The angle of the \bar{B}_E vector, relative to horizontal, is called the dip angle, and the angle between true and magnetic north is called magnetic variation. Tabulated values are available for this field distribution which can be used to correct the heading sensor measured value.

Aircraft incorporating electronic AHRS normally use a magnetometer for measuring \bar{B}_E . Solid state magnetometers based upon Hall, magnetostriction, and Faraday effects are available, although some aircraft use a sensor that is called a flux gate. Figure 20 depicts a typical flux gate configuration. The assembly consists of an outer and an inner sealed chamber, with the intervening space filled with a viscous damping fluid. The outer sealed chamber is attached to the aircraft structure. The inner sealed chamber is mounted on a universal joint such that the excitation coil axis is nominally along the aircraft vertical axis.

The operation of this magnetometer can be understood with reference to Fig. 20. The excitation coil is driven by a 400-Hz sinusoidal current of sufficient strength to saturate the magnetically permeable shell at 800 Hz. During the portion of each cycle that the material is saturated, the

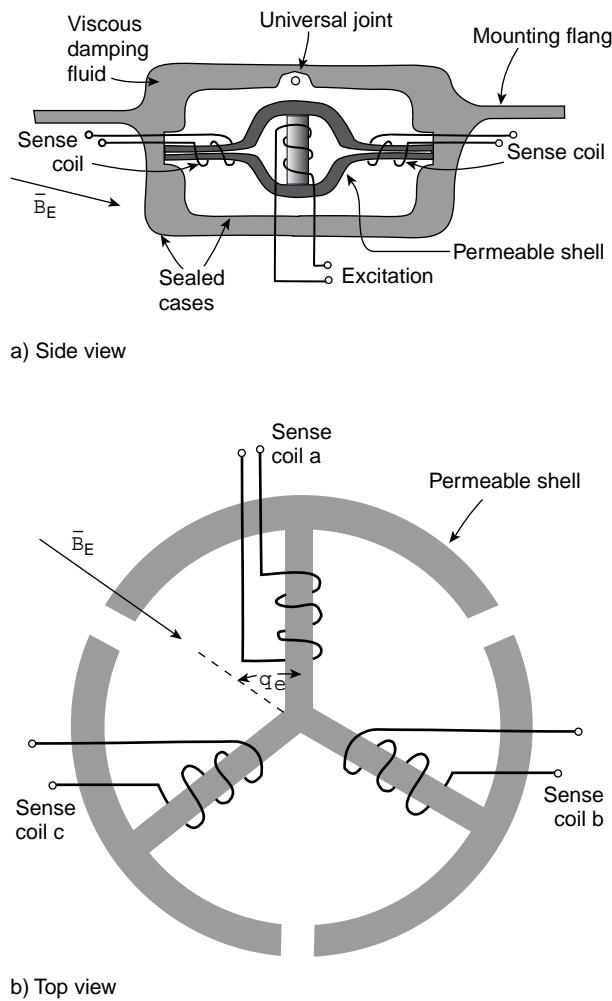


FIGURE 20 Flux gate configuration.

Earth's field passes through the structure without distortion. When the excitation field is zero, the relatively high permeability of this shell distorts \bar{B}_E such that its magnetic flux is coupled to the three sensor coils. A voltage is induced in each sensor coil due to the Earth's field during this portion of the cycle. The amplitude of the induced voltage at 800 Hz is a function of the strength and angle of \bar{B}_E relative to each sensor coil axis. The vector quantity \bar{B}_E can be computed from these three sensor coil voltages. With the proper vertical gyro-erection and the magnetic heading reference from the magnetometer, there is sufficient information for the AHRS to provide attitude and heading measurements.

The AHRS display is either electromechanical or solid state electronic. The electromechanical display for pitch and roll consists of a movable sphere inside an assembly. The sphere is driven such that it maintains one axis aligned

with local vertical. A glass cover is marked with a line that is parallel to the aircraft horizontal plane. The ball pivots about an axis that is orthogonal to the aircraft longitudinal and vertical axis. As the aircraft maneuvers, the reference marking on the glass cover indicates the roll and pitch.

The heading indicator consists of a disk in a plane parallel to the instrument panel that rotates about an orthogonal axis. The disk is marked from 0° to 360° . A glass cover is marked with a pointer along the aircraft vertical axis. When correctly initiated, the displayed value under the pointer gives the aircraft heading relative to magnetic north.

Modern high-performance and transport aircraft utilize a solid state video type display. Symbols representative of traditional AHRS indicators are generated electronically on this display. Such a display is a part of a so-called "glass cockpit" that is discussed later in this chapter.

XX. AUTOPILOT—FLIGHT MANAGEMENT SYSTEMS (FMS)

The operation of an aircraft on a flight is greatly aided by some sort of automatic flight control system. The earliest such system, called an autopilot, had limited functionality but became a robust technology that is relatively inexpensive and is commonly available even on general aviation aircraft. For high-performance general aviation, military, and transport aircraft, the autopilot has evolved to a multimode digital flight control system, called a flight management system (FMS).

The simplest autopilot system has the capability of maintaining attitude, altitude, and/or heading/course in cruise flight only (i.e., it is a form of cruise control). The FMS, on the other hand, can control the aircraft from takeoff through a multisegment flight through a touch-down/landing.

A modern FMS is a digital control system having inputs from attitude, altitude, speed, navigation/position, landing system sensors as well as manual inputs from crew flight controls, and switches/keyboard for selection of operating mode as well as commanded 3-D positions/waypoints and times of arrival at waypoints. The outputs from the FMS drive actuators for pitch, roll, and yaw control as well as thrust. Typically, a flight involves manual takeoff and transition to the first segment of the flight. At some point thereafter, flight control is switched to the FMS which has been preprogrammed with selected waypoints (position and altitude sometimes including desired times of arrival). The FMS regulates the flight path in essentially straight-line segments between successive waypoints. It does this by regulating control surfaces and engine thrust.

During final approach and landing, the crew may either fly manually or enter new waypoints in response to ATC commands. It is not generally possible to preprogram the approach/landing phase since this is a very dynamic circumstance influenced by weather and the sequencing of other arriving aircraft. However, an advanced FMS has the capability of intercepting an ILS localizer and glide slope and executing an automatic landing.

Configuration changes (i.e., flap setting and landing gear extend/retract) normally remain the purview of the flight crew. Also under the control of the flight crew is mode selection. For example, in cruise flight, the crew may select a mode that maintains altitude with extended tolerance such that the aircraft responds to wind gusts more gradually than normal, yielding a “soft ride” while maintaining altitude with a safe clearance from other aircraft.

The configuration of FMS is depicted in the block diagram of Fig. 21.

For transport or high-performance general aviation aircraft, an FMS is normally used in one of many control modes that operate in the neighborhood of some steady flight condition (e.g., constant altitude, constant heading, or tracking an intended course). For military fighter aircraft, flight controls can provide variable closed-loop stability in longitudinal or lateral dynamics as well as course tracking and four-dimensional (4-D) navigational flight control.

For those control modes involving near steady flight, the aircraft dynamics can be adequately represented in the ideal noise-free case by linear dynamics of the form

$$\dot{x} = Ax + Bu$$

$$y = Cx,$$

where x = state vector $\in R^n$ (i.e., an n dimensional vector)

A = state transition matrix

$\in R^{n \times n}$ (i.e., an $n \times n$ matrix)

B = input matrix $\in R^{n \times m}$ (i.e., an $n \times m$ matrix)

u = inputs to the aircraft $\in R^m$ (i.e., an m dimensional vector)

y = output $\in R^\ell$ (i.e., an ℓ dimensional vector)

For any flight condition involving relatively small deviations from a steady state (i.e., a constant state vector x_0), the longitudinal and lateral dynamics are effectively decoupled into longitudinal (i.e., motion in the plane of symmetry) and lateral (i.e., motion orthogonal to the longitudinal).

The state variables for longitudinal dynamics typically include

V = airspeed

α = angle of attack

q = pitch rate

θ = flight path angle (relative to horizontal plane)

The input variables for longitudinal dynamics include

δ_e = elevator deflection

δ_f = flap deflection (including leading edge slats)

T = thrust (assumed symmetric)

The state variables for lateral dynamics include

β = side slip angle

p = roll rate

r = yaw rate

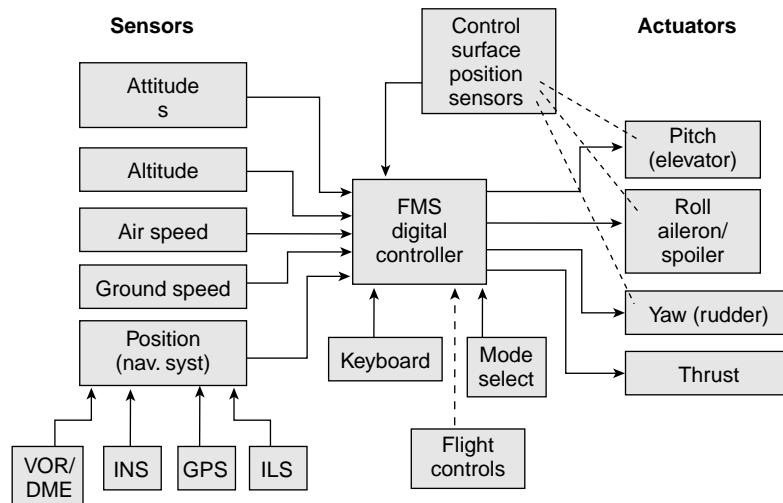


FIGURE 21 Block diagram of representative FMS.

$$\begin{aligned}\phi &= \text{roll angle} \\ \psi &= \text{heading angle}\end{aligned}$$

The inputs for the lateral dynamics include

$$\begin{aligned}\delta_a &= \text{aileron deflection} \\ \delta_r &= \text{rudder deflection} \\ \delta_s &= \text{flight spoiler deflection (asymmetric)}\end{aligned}$$

For flight conditions within the region of validity of the linear dynamic model (i.e., small perturbations about steady flight), the open-loop longitudinal dynamics are characterized by the roots of

$$\det|\lambda I - A| = 0.$$

The roots of this equation are the eigenvalues of the matrix A. Typically, these roots include two complex conjugate pairs in which one is heavily damped and relatively high frequency (called fast mode) and the other is lightly damped and relatively slow (called phugoid mode). For example, the resonant frequency of the phugoid for a large transport aircraft is typically in the range $10^{-2} < \omega < 10^{-1}$ (rad/sec), and the damping ratio is about $0.2 < \zeta < 0.3$. Flight in turbulent air tends to excite this relatively undamped oscillation, which could be quite objectionable to passengers.

Similarly, the lateral dynamics are characterized typically by the five eigenvalues of the lateral dynamics state transition matrix, including at least one complex conjugate pair, two real nonzero roots, and one zero root. The complex pair is associated with a combination roll and yaw oscillation called a "Dutch Roll." In particular, swept wing jet aircraft have such oscillatory motion which is typically very lightly damped and results in an objectionable ride. Normally, at least one of the real roots is slightly positive, indicating an open-loop lateral instability. However, this instability is readily handled by either human pilots or by the FMS/autopilot.

There are many possible control configurations for an FMS, and it is beyond the scope of this article to cover these possibilities. However, we do consider here an example configuration that illustrates the general control theory applicable. In this example, we consider first a separate pitch attitude controller having the block diagram shown in Fig. 22.

The aircraft dynamics are given by

$$\dot{x} = Ax + Bu.$$

The output variables being regulated are y where

$$y = Cx$$

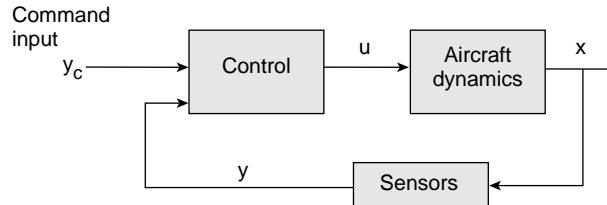


FIGURE 22 Simplified block diagram for the pitch control system.

where C is the matrix associated with the various sensors measuring the state variables. The controller maintains flight at a desired output y_o by generating output u that moves the associated control surface deflections and/or varies engine thrust. The controller generates outputs (i.e., inputs to the aircraft dynamics u):

$$\begin{aligned}u &= K(y_o - y) \\ &= Ke,\end{aligned}$$

where K is the gain matrix of the pitch attitude controller. As long as the error e (where $e = y_o - y$) between desired and actual variable is nonzero, then the control outputs will change the aircraft control surfaces to reduce the error toward zero.

The closed-loop dynamics are characterized by the eigenvalues of the closed-loop error dynamics matrix A_{ce} , where

$$A_{ce} = A - BKC.$$

The eigenvalues of A_{ce} (which are determined by the gain matrix K) can be chosen by the control system designer. By proper choice of the gain matrix K, it is possible, in principle, to vary the damping of the phugoid mode and to suppress it substantially.

The pitch attitude controller can be configured for a variety of modes including altitude hold, speed hold, vertical speed hold, glide path regulation, glide slope tracking, and automatic landing. In implementing many of these modes, it is also necessary to regulate engine thrust to achieve desired performance. Techniques for designing controllers (i.e., selecting the K matrix) are well known in modern control theory and can achieve a variety of optimal performance.

The lateral control system is conceptually similar to the longitudinal controller and can achieve optimized closed-loop performance. One aspect of lateral control that deserves special attention is the suppression of the Dutch roll oscillation. Typically, this is done by means of a special control system that regulates the rudder (or a separately movable portion of it) and is called a yaw damper. A yaw damper is implemented by means of a yaw rate (r) sensor and an actuator that regulates rudder deflection (δ_r) in order to suppress Dutch roll oscillation.

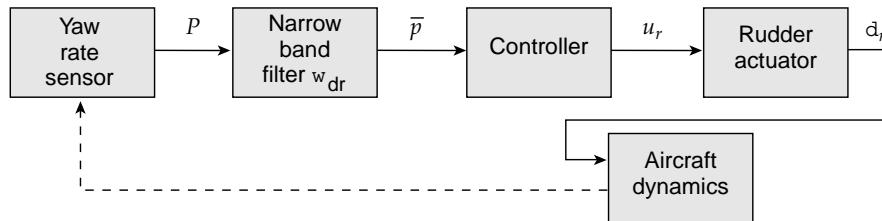


FIGURE 23 Block diagram for a typical yaw damper system.

However, as the rudder is deflected normally as part of a coordinated turn, the rudder deflection under yaw damper authority must only respond to Dutch roll oscillations. Consequently, the yaw damper system typically incorporates a narrow band filter tuned to the Dutch roll frequency ω_{dr} . Figure 23 is a block diagram of a yaw damper.

The aircraft lateral dynamics in this case are given by

$$\dot{x} = Ax + B\delta_r,$$

and

$$\delta_r = -K_{dr}x.$$

By choosing the gain matrix K_{dr} optimally the damping of the Dutch roll mode may be substantially increased from the open loop value, thereby suppressing the Dutch roll oscillation.

In conventional aircraft the flight controls are mechanically connected to the control surfaces. Control forces from the flight crew are augmented hydraulically or electrically for relatively heavy aircraft.

A trend that began with high-performance fighter aircraft and has moved into transport aircraft is to eliminate the mechanical connections from the cockpit controls entirely. A system which is euphemistically called “fly by wire” replaces augmented mechanical controls with an FMS system in which the FMS outputs directly drive the actuators that drive the control surfaces. In such a system, measurements of flight control positions provide inputs to the FMS. Sensors that measure aircraft attitude provide feedback signals for a closed-loop system similar to that shown in Fig. 22.

The FMS provides a multimode, full authority flight control system that yields a very high level of cockpit automation. Fully automatic flight from takeoff through landing is possible with modern fly by wire systems with the flight crew performing a supervisory role.

In military fighter aircraft, stability augmentation that is possible via closed-loop control can convert an aircraft that is open-loop unstable into a stable aircraft. In fact, the stability can be varied dynamically as a function of flight condition in order to maximize combat performance and effectiveness. Typically, such a fighter aircraft is designed

intentionally to be (at least partially) unstable. The variable stability is achieved with a digital FMS in which the feedback gain parameters are changed to vary the closed-loop state transition matrix eigenvalues.

XXI. GLASS COCKPIT

Traditional aircraft instrumentation has separate, stand-alone, cockpit displays for indicating the results of the relevant measurement. That is to say that each variable displayed in the cockpit has a dedicated instrument subsystem with associated display. A recent trend has been to replace traditional instrumentation with multifunction digital instrumentation utilizing a cathode ray tube (CRT) or solid state equivalent reconfigurable display. The latest technology incorporated on transport aircraft uses color flat panel liquid crystal displays (LCD) that have the same display functionality as traditional CRTs. Such an instrumentation system has a block diagram as shown in Fig. 24 or its functional equivalent.

An electronic multifunction instrument system such as that shown in Fig. 24 is euphemistically known as a “glass cockpit.” Normally, a glass cockpit consists of 10 or more individual displays. At least one new 21st century transport utilizes a configuration as shown in Fig. 25.

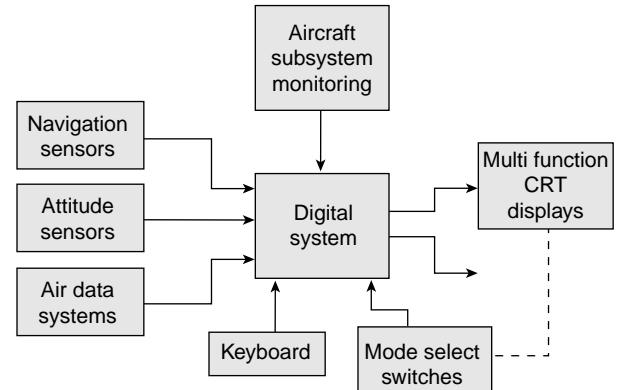


FIGURE 24 Block diagram of an aircraft electronic instrumentation system.

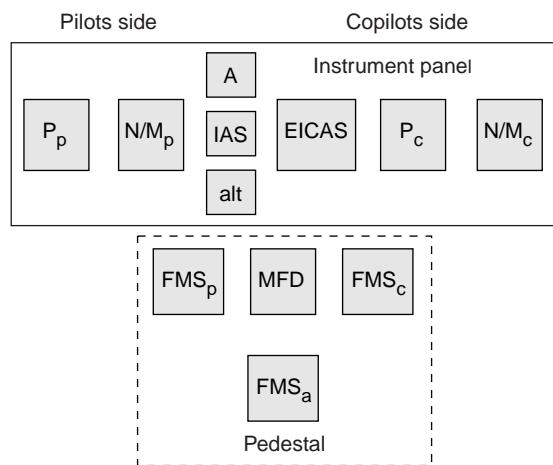


FIGURE 25 Representative glass cockpit configuration.

The displays depicted in this figure include P_p and P_c primary flight displays for pilot (p) and copilot (c); navigation/multifunction displays for pilot and copilot; an engine indicating and crew alerting system (EICAS); a multifunction display that is reconfigurable by the flight crew (MFD); a flight management system displays FMS and standby (backup) instruments for displaying; attitude (A); indicated airspeed (IAS); and altitude (alt).

Each of these display panels is capable of alphanumeric and/or graphic displays as required for its function. The digital subsystem generates signals to drive these displays similar to computer displays using a raster scan format. For example, the attitude display can generate a color pattern that schematically represents the ball and reference horizon of a conventional attitude indicator. At the same time, this display can present numerical data such as airspeed, altitude, heading, rate of climb (normally, after alphanumeric messages from the FMS system are also presented), and descent. Similarly, the navigation display presents a schematic representation of a VOR or HSI indicator in which on selected radial, left/right to/from indications are also depicted.

The engine indicator graphically depicts relevant engine data (e.g., compressor/turbine RPM, turbine inlet temperature, etc.) in a format that resembles conventional engine displays. Use of graphical displays that are suitably configured gives the flight crews the chance to quickly glance at the display to assess that the engine parameters are within the nominal or expected range. Such displays can reduce crew workload during particularly high-task-level flight regimes such as landing in bad weather.

One of the benefits of an integrated digital instrument system is the ability to share data from one system to another. For example, navigation, air data, and FMS data can be combined to yield 4-D navigation (i.e., the ability to fly to a waypoint at a specific altitude and at a specific

time). Furthermore, the existing analytic redundancy in computation of aircraft attitude, navigation data, etc. permits limited compensation for hardware failure such as a sensor failure or display panel failure. In principle, any given glass cockpit display can be reconfigured to perform any desired display function.

This reconfiguration capability is routinely used to permit the flight crew to deal with problems in flight. Selectable synoptic display of aircraft systems or subsystems can be called up to assist in analyzing problems and can replace a pilot's operating manual, which for a transport aircraft is a lengthy, somewhat cumbersome document to use. The glass cockpit along with the high level of automation in the cockpit of a modern transport aircraft has permitted operation with a two-person flight crew and has eliminated the need for a flight engineer (or second officer).

This article on aircraft instrumentation has, by space constraint limits, omitted discussion of many important subsystems. For example, weather radar, collision avoidance, terrain avoidance, and other systems found on today's aircraft could not be included. Moreover, military aircraft which have many avionic systems associated with weapons management and delivery were deemed to be beyond the scope of this article. Nevertheless, the major aircraft instrumentation systems that were discussed here present the reader with a good overview of this interesting subject. Interested readers should consult the bibliography for references to the many subjects that have not been presented here.

Acronyms

AHRS	Attitude Heading and Reference System
AM	Amplitude Modulated
ATC	Air Traffic Control
CDI	Course Deviation Indicator
CRT	Cathode Ray Tube
CSB	Carrier With Sidebands
DG	Directional Gyro
DGPS	Differential GPS
DME	Distance Measuring Equipment
DOD	Department of Defense (USA)
ECEF	Earth Centered-Earth Fixed (Coordinate System)
ECI	Earth Centered Inertial (Coordinate System)
EICAS	Engine Indicating and Crew Alerting System
FAA	Federal Aviation Agency
FADEC	Full Authority Digital Engine Control
FAF	Final Approach Fix
FAR	Federal Air Regulations
FMS	Flight Management System
FSS	Flight Service Station

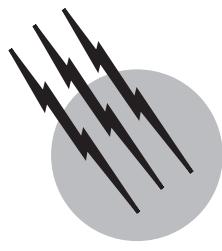
GDOP	Geometric Dilution of Position
GEO	Geosynchronous Earth Orbiting (Satellite)
GLONASS	Global Orbiting Navigation Satellite System
GPS	Global Positioning System
GRI	Group Repetition Interval
GWA	Geodetic Wander Azimuth (Coordinates)
HSI	Horizontal Situation Indicator
IAS	Indicated Airspeed
IF	Intermediate Frequency
IFR	Instrument Flight Rules
ILS	Instrument Landing System
IMC	Instrument Meteorological Conditions
INS	Inertial Navigation System
LAAS	Local Area Augmentation System
LCD	Liquid Crystal Display
LOP	Locus of Points
LORAN	Low-Frequency Radio Navigation System
LPF	Low Pass Filter
MFD	Multifunction Display
MLS	Microwave Landing System
NMI	Nautical Miles
OBS	Omni Bearing Selector
OLS	Ordinary Least Squares
PGTR	Pulse Group Time Reference
RAIM	Receiver Autonomous Integrity Monitoring
RNP	Required Navigation Precision
RPM	Revolutions Per Minute
RVR	Runway Visual Range
SBO	Sideband Only
SSE	Sum Squared Errors
TD	Time Difference
VFR	Visual Flight Rules
VG	Vertical Gyro
VOR	Very High Frequency Omni Range (Navigation System)
WAAS	Wide Area Augmentation System
WMS	WAAS Master Station
WRS	WAAS Reference Station

SEE ALSO THE FOLLOWING ARTICLES

AIRCRAFT AVIONICS • AIRCRAFT PERFORMANCE AND DESIGN • AIRCRAFT SPEED AND ALTITUDE • AIRPLANES, LIGHT • KALMAN FILTERS AND NONLINEAR FILTERS • STOCHASTIC PROCESSES

BIBLIOGRAPHY

- Brown, R. G., and Hwang, P. Y. C. (1992). "Intro to random signals and applied Kalman filtering," 2nd ed., Wiley, New York.
- Campbell, S. P., and LaFrey, R. R. (1984). "Flight test results for an experimental GPS C/A-Code Receiver in a General Aviation Aircraft," Global Positioning System, Vol. 2, pp. 239–257. Institute of Navigation, Washington, DC.
- DeGroot, L. G. (1964). "Navigation and control from Loran-C." *J. Inst. Navigation*, **11** (3).
- Doherty, R. H., Heffrey, G., and Linfield, R. F. (1961). "Timing potentials of Loran-C." *Proc. IRE* **49** (11).
- Durbin, E. (1962). "Current developments in the Loran-C system." *J. Inst. Navigation (U.S.)* **9** (2),
- Enge, P. K., and McCullough, J. R. (1990). "Aiding GPS with calibrated Loran-C." *J. Inst. Navigation* **35** (4), 469–482.
- Enge, P. K. et al., (1990). "Combining pseudo-ranges from GPS and Loran-C for air navigation," *J. Inst. Navigation* **37** (1).
- Kayton, M. (1968). "The near field of an instrument landing system glide slope." *IEEE Trans. Aerosp. Electrical Syst.* **July**, 237.
- Klobuchar, J. A. (1980). "Ionospheric effects on GPS," pp. 48–51, GPS World, Eugene, OR.
- Maybech, P. S. (1979). "Stochastic Models, Estimation and Control," Vol. 1, Academic Press, New York.
- Meditch, J. S. (1969). "Stochastic Optimal Linear Estimation and Control," McGraw-Hill, New York.
- Milliken, R. J., and Zoller, C. J. (1991). "Principle of Operation of NAVSTAR and System Characteristics," Global Positioning System, Vol. 1, pp. 3–14, Institute of Navigation, Washington, DC.
- O'Donnell, C. F., ed. (1964). "Inertial Navigation Analysis and Design," McGraw-Hill. New York.
- Pitman, G. G., ed. (1962). "Inertial guidance," Wiley, New York.
- RTCA, Inc., Washington, DC.
- Samaddar, S. N. (1979). "The theory of Loran-C ground wave propagation—A review." *J. Inst. Navigation* **26** (3).



Aircraft Performance and Design

Francis Joseph Hale

North Carolina State University

- I. Aircraft Flight Behavior
- II. Performance Analysis and Preliminary Design
- III. Aircraft Forces and Subsystems
- IV. Turbojet and Turbofan Aircraft
- V. Piston–Propeller Aircraft
- VI. Turboprop and Turbofan Aircraft
- VII. Conventional Subsonic Aircraft
- VIII. Vertical and Short Takeoff and Landing Aircraft
- IX. Military Aircraft
- X. Advanced Supersonic Commercial Aircraft
- XI. Hypersonic Aircraft and the National Aerospace Plane

GLOSSARY

- Air breathers** Engines that obtain oxygen from the atmosphere.
- Aspect ratio** Wing span divided by average chord length.
- Ceiling** Maximum altitude for specified rate of climb.
- CTOL** Conventional (horizontal) takeoff and landing aircraft.
- Drag polar** Drag coefficient–lift coefficient relationship.
- Flight envelope** Region of possible flight operation.
- HST** Hypersonic transport.
- Lift-to-drag ratio** Aerodynamic efficiency of an aircraft.
- Mass ratio** Aircraft weight at start of cruise divided by aircraft weight at end of cruise.

Operational empty weight Weight of aircraft less the useful load.

Span loading Aircraft weight divided by wing span.

Specific fuel consumption Fuel flow rate per pound of thrust (turbojet) or per horsepower (piston–prop).

Specific impulse Rocket engine thrust divided by fuel flow rate.

SST Supersonic transport.

STOL Aircraft capable of short takeoffs and landings.

Useful load Sum of fuel and payload weights.

V/STOL Aircraft capable of both vertical and short takeoffs and landings.

Weight fraction Component weight divided by aircraft weight.

Wing loading Aircraft weight divided by wing area.



FIGURE 1 The Boeing 777. A long-range (8200 mi), wide-body, Mach 0.85 transport with 300 passengers, two high-bypass-ratio turbofans of the 90,000-lb thrust class, a maximum gross weight of 630,000 lb, a wing area of 4600 ft² (W/S = 137 lb/ft²), and an aspect ratio (AR) of 8.7. [Courtesy of Boeing.]

AN AIRCRAFT performance analysis starts with the physical characteristics of the aircraft, either actual or hypothetical, and ends with knowledge as to how the aircraft might fly (e.g., its range, payload, airspeeds, ceiling, rate of climb, and turning rate). It does not necessarily concern itself with determining whether the aircraft is flyable and controllable. The converse of the performance analysis is the design of an aircraft to meet a set of operational requirements and constraints. Design, however, is an iterative process, and the first tries, which should yield major aircraft characteristics, are often referred to as conceptual or feasibility design. Many assumptions and approximations are used in early performance and design iterations to give qualitative answers with a minimum expenditure of time and effort. (Figures 1–3 show a Mach 0.85 transport, an SST, and a supersonic fighter, respectively.)

In spite of the push for metric (SI) units, the units used for aircraft calculations, cockpit instruments, and operations are, in general, the customary English units, with knots (nmi per hour) being the operational measure of airspeed. This article uses such units but with miles per hour (mph) or feet per second (fps) as the measure of airspeed.

I. AIRCRAFT FLIGHT BEHAVIOR

The flight path and behavior of an aircraft are determined by the interaction between the characteristics of the air-

craft and the environment in which it is flying. The aircraft characteristics can be categorized as the physical characteristics, such as the shape, mass, volume, and surface area; the characteristics of the subsystems, such as the propulsion, guidance, and control systems; and the structural characteristics, such as loading and temperature limitations and the stiffness (rigidity) of the structure.

The environment affects the flight of an aircraft through the field forces and the surface forces. The only field force that need be considered is gravity, which appears as the weight and is a function of the mass. The surface forces are the aerodynamic forces (the lift, the drag, and the side force) which are strongly dependent on the shape and surface area of the aircraft, especially of the wings; the properties of the atmosphere; and the airspeed. In addition, consideration must be given to the inertial forces, which are a consequence of nonequilibrium processes and maneuvers and which play an important role in the dynamic and stress analyses of aircraft but not in performance analyses; inertial forces are ignored in this article.

Although most modern aircraft have various amounts of elasticity arising from the desire to reduce the structural weight to a minimum, aeroelastic effects need not be considered in initial performance analyses and preliminary configuration designs. Furthermore, the three rotational modes of a rigid aircraft can be ignored by considering the aircraft to be a point mass, and the inertia forces can



FIGURE 2 The Concorde SST, the only civil supersonic transport (in service for over 30 years). A medium-range, Mach 2.0 transport with 100 passengers, four afterburning turbojets of 38,000-lb unaugmented thrust, a maximum gross weight of 408,000 lb, a wing area of 4134 ft^2 ($W/S = 98.6 \text{ lb}/\text{ft}^2$), and an AR of 1.7. [Courtesy of British Aerospace.]

be neglected (with the exception of the centrifugal force in turning flight) by assuming that velocities and other flight-path parameters are either constant or changing so slowly that their rates of change can be neglected. However, an

acceleration correction should be made for the climb of a high-performance aircraft.

With the assumptions and approximations of the preceding paragraphs, the performance problem can be



FIGURE 3 The F-22 Raptor, a joint Lockheed Martin/Boeing aircraft. A single-place, Mach 2, air-dominance fighter with two low-bypass turbofans of the 35,000-lb class with supercruise capability, a wing area of 840 ft^2 , and an AR of approximately 2.3. [Courtesy of Boeing.]

reduced to a set of simple, two-dimensional statics problems that, although idealized and not applicable to all classes of aircraft or flight regimes, are amazingly powerful in providing insight into the reasons that conventional aircraft look and fly as they do. It is interesting to discover that turbojet aircraft fly high and fast not necessarily because they want to but because they must do so if they are to be competitive with piston-propeller aircraft.

II. PERFORMANCE ANALYSIS AND PRELIMINARY DESIGN

A performance analysis takes an existing set of characteristics of a particular aircraft and determines such things as how fast and how high it can fly and how far it can travel with a specified amount of fuel and payload. Performance also means examining different ways of flying a mission so as to exploit the characteristics and capabilities of the aircraft in question. For example, two major objectives of commercial aircraft are to minimize both the fuel consumption and the flight time, objectives that can be met by a judicious combination of proper design and appropriate operational procedures.

The aircraft undergoing a performance analysis may be a real aircraft or it may be a paper (hypothetical) aircraft that is being studied for possible adoption and manufacture. As a result of the performance analysis, it may be rejected or some of its characteristics may be modified and another performance analysis carried out. Thus, the performance analysis and the preliminary design process are often intertwined in a series of iterations, and it may be difficult at times to distinguish between the two. If it is possible to determine the operational performance of an aircraft with a set of specified characteristics, it seems logical to be able to do the converse, that is, to design an aircraft that will satisfy a set of operational requirements, such as desired range, cruise airspeed, and payload capability.

There are three basic approaches to a performance analysis: graphic, analytical, and computational. Although all three have advantages and proponents, the analytical approach is favored in this article for preliminary performance and design analyses since it permits classifying and describing aircraft on a broader basis than does the graphic approach and may give greater insight into physical interactions than does the computational approach, which is essentially a numbers in-numbers out process. With the analytical approach a knowledge of the wing loading, the drag polar, and the type of propulsion will tell us much about the design mission and performance of an aircraft.

With an awareness of the importance and significance of certain aerodynamic and propulsive parameters and of the overriding importance of structural weight and integrity in

the design of an aircraft and the impact on initial and operating costs, a person need not have a detailed background in these technical areas in order to perform preliminary performance and design studies.

III. AIRCRAFT FORCES AND SUBSYSTEMS

A. The Atmosphere

Since the thrust and aerodynamic forces of an aircraft are strongly influenced by the properties of the atmosphere, some knowledge of these properties is essential. The International Standard Atmosphere is based on the Air Research and Development Command (ARDC) Model Atmosphere of 1959, which has seven concentric layers surrounding the earth. The layer next to the surface of the earth (starting at sea level) is called the *troposphere* and is characterized by a decreasing ambient temperature. At 36,089 ft above mean sea level, the temperature remains constant up to an altitude of 82,021 ft above mean sea level. This second (isothermal) layer is called the *stratosphere*, and the separating altitude of 36,089 ft is known as the *tropopause*. The sea-level properties for a standard day are listed in [Table I](#) along with the ratios of the properties at a given altitude to the sea-level values.

TABLE I Standard Atmosphere Property Ratio^a

Altitude (10^3 ft)	Density ratio $\sigma = \rho / \rho_{SL}$	Speed of sound ratio $a^* = a / a_{SL}$
0	1.000	1.000
5	0.862	0.983
10	0.738	0.965
15	0.629	0.947
20	0.533	0.929
25	0.448	0.910
30	0.374	0.891
35	0.310	0.871
36	0.297	0.867
40	0.246	0.867
45	0.194	0.867
50	0.152	0.867
55	0.120	0.867
60	0.094	0.867
65	0.074	0.867
70	0.058	0.867
75	0.046	0.867
80	0.036	0.867
82	0.033	0.867
85	0.028	0.872

^a Sea-level values: $P = 2116 \text{ lb}/\text{ft}^2$; $\rho = 23.769 \times 10^{-4} \text{ lb} \times \text{sec}^2/\text{ft}^4$; $a = 1116 \text{ fps}$; $R = 1.7165 \times 10^3 \text{ ft}^2/\text{sec}^2 \times {}^\circ\text{R}$.

Although the values from the tables are more accurate and normally should be used, there are times when it is desirable to have a mathematical expression for the *density ratio* σ in terms of the altitude. A commonly used expression is

$$\sigma = \rho_1 / \rho_{SL} = \exp(-h_1 / \beta) \quad (1)$$

where ρ_1 is the atmospheric density at altitude h_1 on a standard day, the subscript SL denotes sea-level values, and β is a factor that can have one of two values. A value of 30,500 ft gives good results in the troposphere but introduces large errors at higher altitudes. For altitudes above the tropopause and up to $\sim 250,000$ ft, a value of 23,800 ft gives better results.

Returning to the sea-level properties for a standard day, the atmosphere can be assumed to satisfy the equation of state for an ideal gas, namely,

$$P = \rho RT \quad \text{or} \quad \rho = \frac{P}{RT} \quad (2)$$

where P is the pressure, R the gas constant for air, and T the air temperature expressed as an absolute temperature (Rankine or Kelvin). Consequently, an increase in the ambient temperature over the standard will decrease the density, as will a decrease in the pressure. In other words, on a hot day and/or with a below-standard barometric pressure, the sea-level density will be lower than standard and the performance of the aircraft will be affected. With a reduced atmospheric density, the aircraft and propulsion system think and perform as though they were at a higher altitude than they actually are. Although a standard day is normally assumed, one should be aware of the possible effects of any deviations.

B. Aerodynamic Forces and Considerations

The resultant aerodynamic force produced by the motion of the aircraft through the atmosphere can be resolved into three components along the wind axes with the x axis always along the velocity vector. The component along the x axis is the drag, has the symbol D , and is in the opposite direction to the airspeed and resists the motion of the aircraft. The component perpendicular to the aircraft velocity is the lift, has the symbol L , and functions mainly to counteract the weight of the aircraft (keep the aircraft in the air). The third component, the side force, which appears when there is a sideslip angle (uncoordinated flight), is an undesirable condition and will be neglected. Only the lift and drag forces are considered in this article.

With a well-designed aircraft, the wing is the major source of lift and drag; the design objective is to maximize the lift while minimizing the drag. Wing parameters of importance with respect to performance are the wing span

b , the distance from wing tip to wing tip; the average chord \bar{c} , the average distance from the front (leading edge) of the wing to the back (trailing edge); and the wing area S , the area of one side of the wing to include the area occupied by the fuselage. The ratio of the wing span to the average chord is the *aspect ratio* AR , which is a measure of the narrowness of the wing. The following relationships are useful:

$$S = b\bar{c}; \quad AR = b/\bar{c} = b^2/S \quad (3)$$

The thickness ratio t/\bar{c} is of importance in determining wing drag. The cross section of a wing is called an airfoil and can be symmetrical or asymmetrical. Many current airfoils for conventional aircraft are symmetrical or nearly so.

The lift and drag forces can be obtained from the following expressions:

$$L = qSC_L = \frac{1}{2}\rho V^2 SC_L \quad (4a)$$

$$D = qSC_D = \frac{1}{2}\rho V^2 SC_D \quad (4b)$$

In these equations, q is the dynamic pressure, has dimensions of pounds per square foot, and is equal to $\frac{1}{2}\rho V^2$, where ρ , the atmospheric density ($\text{lb} \cdot \text{sec}^2/\text{ft}^4$), can be expressed as $\rho_{SL}\sigma$, V is the true airspeed (fps), and S is the wing area. C_L and C_D are the dimensionless lift and drag coefficients, respectively, and are functions of the angle of attack, the Mach number, the Reynolds number, and the shape of the wing; C_L increases with increasing angle of attack until a maximum value $C_{L\max}$ occurs at the stall and then decreases, usually sharply. Conventional aircraft do not fly beyond the stall point. T is the thrust of the engine in pounds and is assumed to be collinear with the airspeed.

The drag polar is an expression of the drag coefficient as a function of C_L rather than of the angle of attack and is very important for performance analyses. One generalized form of the drag polar is

$$C_D = C_{D0} + KC_L^x \quad (5)$$

where C_{D0} is the *zero-lift drag coefficient* (the drag coefficient when the lift coefficient is equal to zero) and the second term on the right-hand side represents the *drag-due-to-lift coefficient*; C_{D0} , K , and the exponent x are all functions not only of the wing shape but also of the Mach and Reynolds numbers. Fortunately, there are many efficient subsonic and thin-winged supersonic configurations and flight regimes where these parameters can be assumed constant and x can be set equal to 2. Consequently, Eq. (5) can be written as the *parabolic drag polar*,

$$C_D = C_{D0} + KC_L^2 \quad (6)$$

The value of K is strongly dependent on the aspect ratio, as can be seen from the expression:

$$K = 1/\pi AR e \quad (7)$$

where e is the *Oswald span efficiency*. An ideal wing with an infinite span would have a value of unity for e . Although practical values of e range from 0.6–0.9 (generally decreasing with a decrease in the aspect ratio), the use of winglets, tip tanks, and end plates increases the span efficiency. With the assumption that C_{D0} is the minimum-drag coefficient of the entire aircraft, only the parabolic drag polar is used in this article. Typical subsonic values of C_{D0} are of the order of 0.011 for a clean fighter aircraft at subsonic airspeeds, 0.016 for a jet transport, and 0.025 for a propeller-driven aircraft.

A very important performance parameter is the *lift-to-drag ratio* (L/D) of the aircraft, which is also the aerodynamic efficiency. It is given the symbol E and defined as L/D or C_L/C_D and for a given shape is a function of the lift coefficient itself and thus of the airspeed. For a parabolic drag polar it can be shown that the maximum value of L/D (E_m) occurs when the aircraft is flying so that the zero-lift drag and drag-due-to-lift coefficients are equal, which occurs when

$$C_{L,E_m} = (C_{D0}/K)^{1/2} \quad (8a)$$

and the maximum lift-to-drag ratio is

$$E_m = 1/2(KC_{D0})^{1/2} \quad (8b)$$

Example 1. An aircraft has a wing span of 190 ft, an average chord of 23.75 ft, a $C_{D0} = 0.016$, and an $e = 0.85$.

1. Write an expression for the parabolic drag polar:

$$AR = 190/23.75 = 8$$

$$K = 1/[\pi(8)(0.85)] = 0.0468$$

$$C_D = 0.016 + 0.0468C_L^2$$

2. Find E_m and the corresponding C_L :

$$E_m = \frac{1}{2(0.0468 \times 0.016)^{1/2}} = 18.27;$$

$$C_L = \left(\frac{0.016}{0.0468} \right)^{1/2} = 0.585$$

3. Find C_D and E for a C_L of 0.3:

$$C_D = 0.016 + 0.0468(0.3)^2 = 0.020$$

$$E = 0.3/0.02 = 15$$

The maximum lift-to-drag ratio E_m is a design characteristic of a particular aircraft because K and C_{D0} are design, not flight, parameters. Each aircraft has its own value of E_m that it cannot exceed, although it can fly at lower values. Some order-of-magnitude values of E_m for several classes of aircraft are the following:

Sailplanes	35
M 0.8 transports	18
Subsonic fighters	10
Supersonic aircraft	7
Helicopters	3

Since the maximum range of a flight vehicle for a given fuel load is directly proportional to the value of E_m , it should be apparent as to why there are no transcontinental helicopters as yet and why there are those who doubt the economic soundness of an SST.

The parabolic drag polar, with constant C_{D0} and K , can be considered valid for airspeeds up to those approaching the drag-rise Mach number M_{DR} , at which point the drag starts to rise (due to the formation of shock waves in the transonic region) and E_m begins to fall off. Techniques for increasing M_{DR} include wing sweep, reduction of the thickness ratio and/or aspect ratio of the wing, a supercritical airfoil, and application of the equivalent area rule.

The shape and size of the wings of an aircraft often indicate the speed, range, and purpose for which the aircraft has been designed. Large, straight, wide (where *wide* refers to the chord length), and thick wings (often cambered) indicate a low-speed, short-range aircraft with emphasis on low stall speeds and short takeoff and landing runs. As the speed (and range) of the aircraft increase, the wings become smaller, narrower, and thinner, and the camber (asymmetry) decreases. There are trade-offs with respect to the wing thickness, however. A thinner wing has lower drag and a higher drag-rise Mach number but at the cost of a higher weight and reduced volume (for fuel, flap mechanisms, or landing gear storage).

At a Mach number of the order of M 0.7 wing sweep begins to appear, reaching an angle of the order of 35° for a M 0.85 subsonic transport or bomber. Although aircraft often fly beyond the drag-rise Mach number, no aircraft has yet been designed to operate in the transonic region and so the next increase in airspeed results in a supersonic aircraft. The wings become thinner, the leading edges are much sharper, the sweep is more pronounced, and the aspect ratio begins to decrease (the wings become stubbier). The combination of increased wing sweep and decreased aspect ratio often results in a delta wing. There are several aircraft with variable-sweep wings so as to tailor the wing configuration to the airspeed at which the aircraft is operating.

C. Propulsion Subsystems

Although a knowledge of the inner workings of an aircraft engine is not necessary, the functional relationships for the propulsive force (the thrust T) and the fuel consumption

rate, along with a feel for the weight of the engine per pound of thrust or horsepower, are required for performance analyses and preliminary design studies. The four types of aircraft engines in current use are all classified as air breathers in that they use oxygen from the atmosphere to burn with a petroleum product fuel, either gasoline or a form of kerosene (commonly referred to as jet fuel). These four types can be further divided into those without propellers, namely, the pure *turbojet* and the *turbofan*, and those with propellers, namely, the *piston-prop* (the reciprocating engine and propeller combination) and the *turboprop*. As will be shown, the flight characteristics of an aircraft are strongly affected by the presence or absence of a propeller. In the following sections, it is assumed that the engines have been properly sized for the aircraft and flight regimes of interest.

The *turbojet* engine produces thrust by expanding hot combustion gases through a nozzle. This thrust, to a first approximation, can be considered to be independent of the airspeed and, for a given throttle (percent rpm) setting, to be directly proportional to the atmospheric density, so that:

$$T_1/T_{SL} = \rho/\rho_{SL} = \sigma_1 \quad (9)$$

Equation (9) shows that the thrust will be at a maximum at sea level and decrease with increasing altitude.

The fuel consumption rate is described in terms of the *thrust specific fuel consumption* (tsfc), with the symbol c , and defined as the fuel weight flow rate per hour per pound of thrust with the units of pounds per hour per pound, usually expressed as inverse hours (hr^{-1}):

$$\text{tsfc} = c = \frac{dW_f/dt}{T} \quad (10)$$

The specific fuel consumption is an engine characteristic and is assumed to be constant for all flight conditions even though it is a function of the airspeed, throttle setting, and altitude. The tsfc is less affected by the altitude than is the thrust. Within the troposphere it decreases as the 0.2 power of the density ratio, reaching a minimum at the tropopause and then increasing very slowly within the stratosphere.

The thrust produced by turbojet engines ranges from 50 lb up to the order of 50,000 lb. The uninstalled thrust-to-engine-weight is constantly increasing and currently is on the order of 4 to 6 lb of thrust for every pound of engine weight. Turbojet engines (actually very low-bypass turbofans) are primarily used in supersonic aircraft and usually have an *afterburner* that burns added fuel with the excess (unburned) oxygen in the gaseous mixture, leaving the turbine to produce additional thrust. Although the thrust is approximately doubled with an afterburner, this *thrust augmentation* is accompanied by a large increase in the tsfc (on the order of two to three times). Consequently, an afterburner is installed only when the operational require-

ments demand one and then is used sparingly, such as for takeoff and climb, for attaining supersonic airspeed, and for periods of supersonic flight. When supersonic flight can be achieved and maintained without using an afterburner, it is called *supercruise*.

The *piston-prop* is an internal combustion engine that burns air and gasoline (there are no diesel aviation engines yet) and produces shaft power rather than thrust. The power output is commonly measured in units of horsepower (hp), is essentially independent of the airspeed, and is a function of the altitude and throttle setting. The fuel consumption rate is proportional to the horsepower (HP), so that:

$$dW_f/dt = \hat{c}\text{HP} \quad (11)$$

where \hat{c} is the horsepower specific fuel consumption (hpsfc) with units of pounds per hour per horsepower.

The propeller converts the engine shaft power into the thrust power P , which is equal to the product of the thrust and the airspeed, where the thrust is in pounds and the airspeed is expressed in feet per second, miles per hour, or knots, whichever is more convenient at the time. The engine horsepower and the propeller thrust power are related by the expression:

$$P = TV = k\eta_p\text{HP} \quad (12)$$

where η_p is the propeller efficiency (of the order of 80–85% for a constant-speed propeller) and k is a conversion factor with a value of 375 when V is in miles per hour and 550 when V is expressed in feet per second. Notice that, for a given horsepower, the thrust power is independent of the airspeed but the available thrust is inversely proportional to the airspeed, decreasing as the airspeed increases, whereas the thrust of a turbojet is constant and the thrust power increases as the airspeed increases. It is these differences that are primarily responsible for the fact that propeller aircraft and jet aircraft fly quite differently for best performance.

The hpsfc has the same variations with altitude as the tsfc and will be assumed to be constant. The variation of the power with altitude depends on whether the engine is supercharged. If not supercharged, the engine is referred to as aspirated, and the power to a first approximation is directly proportional to the atmospheric density, as is the thrust of the turbojet. The superchargers of today use a turbine driven by the engine exhaust gases to increase the density of the air entering the cylinders and are called turbochargers. With a constant throttle setting, the output power remains constant up to the critical altitude, which has a maximum value of the order of 20,000 ft. Above the critical altitude, the power of a turbocharged engine decreases with altitude in the same manner as an aspirated engine.

Current piston-prop engines are relatively small (ranging in size from ~ 50 hp to the order of 600 hp) because they are the heaviest of all the engines. The uninstalled horsepower-to-engine-weight ratio is of the order of 0.8 hp/lb of engine weight.

Turboprop engines and turbofan engines are basically turbojet engines in which the combustion gases are more fully expanded in the turbine section so as to develop more power than is needed to drive the compressor and accessories. This excess power is then used to drive either a propeller, in the case of a turboprop, or a multibladed ducted fan, in the case of a turbofan. Any energy remaining in the gaseous mixture leaving the drive turbines is then expanded in a nozzle to produce what is known as jet thrust. This jet thrust is obviously considerably less than that produced by a comparable turbojet, and in the case of a turboshaft engine is zero.

In the *turboprop* engine, the residual jet thrust is converted to an equivalent horsepower at some design airspeed and the engine is then described in the terminology of the piston-prop, using an equivalent shaft horsepower (eshp) and an equivalent shaft horsepower specific fuel consumption. Since the jet thrust power is of the order of 20% or less of the total power, it is reasonable to treat the turboprop as though it were an aspirated piston-prop. Derated turboprops, which are operated at less than the maximum eshp, have the characteristics of a turbocharged piston-prop and are becoming popular. The turboprop has a slightly higher specific fuel consumption than does the piston-prop, but the weight of the engine is considerably less, even with the weight of the propeller gearbox. Horsepower-to-engine-weight ratios are of the order of 2 hp per pound of engine weight, and the largest engine in current operational use is of the order of 6000 eshp.

Although a *turbofan* engine is described as though it were a turbojet, its characteristics are determined by the bypass ratio, which is the ratio of the mass of the "cold air" passing through the fan to the mass of the "hot air" passing through the burners and turbine section. If the bypass ratio is zero, the turbofan is obviously a pure turbojet. As the bypass ratio increases, the percentage of jet thrust decreases and the turbofan begins to take on the characteristics of a turboprop. For example, with a bypass ratio of 10, the theoretical jet thrust would be of the order of 17%. Current maximum bypass ratios are of the order of 5–6 and tsfc's, for a specified airspeed, are of the order of 0.6 lb/(hr · lb). Although the frontal area of a turbofan increases rapidly with an increase in the bypass ratio, the length decreases; consequently, there is a smaller increase in the engine drag and weight than might be expected. Thrust-to-engine-weight ratios are of the order of 5–6 and increasing, as is the maximum thrust of an individual engine, currently of the order of 98,000 lb.

The specific fuel consumption is an extremely important performance parameter. Some typical values, all expressed as an equivalent tsfc (lb/hr · lb), are

Rocket engines	10
Ramjets	3
Turbojets (afterburner)	2.5
Turbojets	0.9–1
High-bypass turbofans	0.6–0.8
Turboprops	0.5–0.6
Piston-props	0.4–0.5

It is interesting that, for various reasons, this list also indicates the relative airspeed regime of the flight vehicles in which these engines are used. For example, piston-prop engines are used in aircraft with airspeeds of the order of 250 mph or less; turboprop engines, at the higher airspeeds up to approximately M 0.7; turbofan engines, for airspeeds up to M 0.85; and turbojet engines and very low bypass engines, in supersonic aircraft. The ramjet is suitable for flight vehicles at M 3.0 and higher, and rocket engines are used in ballistic missiles and space boosters. Furthermore, the piston-prop engine is the least expensive and the heaviest of the engines, the weight decreasing and the cost increasing as the list is ascended.

D. Weight Fractions

Weight (along with the propulsion system) may well be the most important consideration in the design and performance of an aircraft. Design experience has shown that often the lowest weight design is also the lowest cost and most efficient design. Every extra pound of weight is accompanied by an increase in the wing area, thrust, fuel, and so on, all leading to a further increase in the aircraft weight and adversely affecting the performance and costs (both initial and operational) of the aircraft. Weight fractions are very useful in performance and design analyses and are obtained by expressing the gross (total) weight of an aircraft as the sum of the individual components and subsystems and then dividing by the gross weight.

Although the "bookkeeping" process can be, and in final design is, very detailed, the weight of the aircraft can be considered to be made up of the structural weight W_s , the engine weight W_e , the payload weight W_{PL} , and the fuel weight W_f . The structural weight, with this breakdown, includes not only the weight of the structure itself but also the weight of everything not included in the other categories. It includes the weight of all the equipment and landing gear, for example, and even the weight of the flight and cabin crews. Aircraft manufacturers often lump the structural and empty weights together and call the sum the empty weight or the operational empty weight and combine the fuel and payload weights into the useful load.

With these divisions, the gross weight of the aircraft W_0 can be written:

$$W_0 = W_s + W_e + W_{PL} + W_f \quad (13)$$

Dividing through by the gross weight yields:

$$1 = \frac{W_s}{W_0} + \frac{W_e}{W_0} + \frac{W_{PL}}{W_0} + \frac{W_f}{W_0} \quad (14)$$

Equation (14) defines the weight fractions: the structural weight fraction, the engine weight fraction, the payload weight fraction, and the fuel weight fraction. The sum of these individual weight fractions must always be unity. The engine weight fraction for turbojets and turbofans is equal to the ratio of maximum engine thrust to the gross weight of the aircraft divided by the ratio of the maximum engine thrust to the engine weight. A similar relationship, using horsepower rather than thrust, is used for propeller aircraft.

The analysis and design of aircraft structures and the determination of component and subsystem weights are complex and beyond the scope of this article. Order-of-magnitude values for the structural weight fraction will be used instead in order to give a feel for its size and significance. As the gross weight of an aircraft increases, its structural weight fraction decreases. There are minimum volume requirements for cabin and cargo space, and the fixed weights of equipment (such as avionics and navigation systems) have a larger impact on the structural weight fractions of the lower gross weight aircraft. Large subsonic transports, such as the C-5 and 747, have structural weight fractions of the order of 0.4–0.45, whereas smaller aircraft, such as general aviation and fighter aircraft, can have values in excess of 0.5. Remember that this structural weight fraction does not include the weight of the engines.

One should never forget nor underestimate the importance and significance of the weight fractions. For example, a payload weight fraction of 0.1 means that 10 lb of aircraft weight is required for each pound of payload (or excess weight). Therefore, a payload of 2000 lb calls for a 20,000-lb aircraft. The weight fractions can be used to demonstrate dramatically the range and payload limitations associated with the increased engine weight required for a true vertical takeoff and landing (VTOL) aircraft as compared with a comparable conventional takeoff and landing (CTOL) aircraft.

Example 2. An aircraft is to have a payload capability of 10,000 lb using engines with a T_m/W_e of 5 and an estimated W_s/W_0 of 0.45. A performance analysis calls for a W_f/W_0 of 0.3.

- For a CTOL aircraft with a T_m/W_0 of 0.25, what is the estimated gross weight?

$$W_e/W_0 = 0.25/5 = 0.05$$

$$0.45 + 0.05 + W_{PL}/W_0 + 0.3 = 1$$

$$W_{PL}/W_0 = 0.2$$

$$W_0 = 10,000/0.2 = 50,000 \text{ lb}$$

- For a VTOL aircraft with a T_m/W_0 of 1.1, what is the estimated gross weight?

$$W_e/W_0 = 1.1/5 = 0.22$$

$$W_{PL}/W_0 = 1 - (0.45 + 0.22 + 0.3) = 0.03$$

$$W_0 = 10,000/0.03 = 333,333 \text{ lb}$$

- If the gross weight of the VTOL of question 2 is held at 50,000 lb, what is the fuel weight fraction?

$$W_{PL}/W_0 = 10,000/50,000 = 0.2$$

$$W_f/W_0 = 1 - (0.45 + 0.22 + 0.2) = 0.13$$

IV. TURBOJET AND TURBOFAN AIRCRAFT

A. Cruise Flight

For any aircraft in unaccelerated level flight, the lift must be equal to the weight if the aircraft is to remain at its altitude. The lift is developed by moving the wing through the air. This movement is resisted by the drag, which must in turn be balanced by the thrust if the airspeed is to remain constant. The weight, however, is not constant, decreasing as fuel is used, and the rate at which the aircraft weight is decreasing is the negative of the fuel consumption rate, which is determined by the type of propulsion used. Consequently, the propulsion system must be specified before a performance analysis can be executed.

Since the turbojet engine is a thrust producer and since the thrust appears explicitly in the equations of motion, the performance analysis of a turbojet is relatively straightforward. Furthermore, the results can be applied directly to the turbofan, which can also be treated as a thrust producer, and can be useful in evaluating the performance of propeller aircraft, namely, the piston-prop and turboprop. The relevant equations for a turbojet/turbofan are

$$T = D; \quad L = W \quad (15a)$$

$$dX/dt = V; \quad dW/dt = -cT \quad (15b)$$

In these equations, with the assumptions made earlier, the thrust is a function of the altitude and throttle setting, and the drag (since the lift is equal to the weight) is a function of the altitude, the true airspeed, and the weight. Although the V in the first expression of Eq. (15b) is actually the ground speed, it is customary to assume that there is no

wind (the still-air assumption) so that the airspeed and ground speed are identical.

With the introduction of the parabolic drag polar, the drag can be expressed as:

$$D = qSC_D = qSC_{D0} + qSKC_L^2 \quad (16)$$

where q is the dynamic pressure and is equal to $\frac{1}{2}\rho_{SL}\sigma V^2$. Whether q or its equivalent expression is used is a matter of convenience. With $L = W$, Eq. (4a) can be solved for C_L to obtain:

$$C_L = \frac{W/S}{q} = \frac{2(W/S)}{\rho_{SL}\sigma V^2} \quad (17)$$

where W/S is called the wing loading of the aircraft, has the units of pounds per square foot, and is an important performance and design parameter. Substituting Eq. (17) into Eq. (16) yields two drag expressions that are very useful. One expression is for the drag as a function of the weight, and the other (which is even more useful) is for the drag-to-weight ratio (which is also the level-flight thrust-to-weight ratio) in terms of the wing loading. These expressions are

$$D = qSC_{D0} + \frac{KW^2}{qS} = \frac{1}{2}\rho_{SL}\sigma V^2 SC_{D0} + \frac{2KW^2}{\rho_{SL}\sigma V^2} \quad (18a)$$

$$\frac{D}{W} = \frac{qC_{D0}}{(W/S)} + \frac{K(W/S)}{q} = \frac{\rho_{SL}\sigma V^2 C_{D0}}{2(W/S)} + \frac{2K(W/S)}{\rho_{SL}\sigma V^2} \quad (18b)$$

Since the thrust must be equal to the drag in level flight, either of these expressions can be solved for the level-flight airspeed, which is found to be

$$V = \left\{ \frac{T/S}{\rho_{SL}\sigma C_{D0}} \left[1 \pm \left(1 - \frac{4KC_{D0}}{(T/W)^2} \right)^{1/2} \right] \right\}^{1/2} \quad (19)$$

The \pm sign in Eq. (19) indicates that mathematically there are two possible values for the level-flight airspeed for a given aircraft with a specified throttle setting, altitude, and weight, values that may not necessarily be physically realizable.

In Eq. (19) the radicand $1 - [4KC_{D0}/(T/W)^2]$ must be equal to or greater than zero if there are to be any real solutions. With the introduction of the E_m relationship of Eq. (18b), the necessary condition for steady-state level flight is

$$T/W \geq 1/E_m \quad (20)$$

where T is the actual, instantaneous thrust, which is a function of the altitude and the throttle setting, and W

is the actual, instantaneous gross weight of the aircraft. Remember that E_m (the maximum lift-to-drag ratio) is a design characteristic of the aircraft.

Figure 4 is a graph of typical level-flight equilibrium conditions for a given altitude and wing loading. For a given throttle setting the available thrust-to-weight ratio (T_1/W) is, to a first approximation, independent of the airspeed. The drag-to-weight ratio, however, is strongly dependent on the airspeed, being very large at low airspeeds (where the drag due to lift is dominant, since the lift coefficient is large), decreasing with increasing airspeed to reach a minimum value at E_m , and then becoming very large again at high speeds (where the zero-lift drag is dominant). For the throttle setting for T_1/W , there are two points of intersection of the two curves, points where the thrust is equal to the drag. Point 1 (V_1), is the high-speed solution and point 2 (V_2) is the low-speed solution.

It is important to remember that V_1 and V_2 may not be physically realizable— V_1 if it exceeds the drag-rise Mach number, and V_2 if it is less than the stall speed. Furthermore, it can be shown that V_2 is a statically unstable condition in that, if the airspeed decreases for some reason to V_5 , the thrust will be less than the drag and the aircraft will decelerate rapidly. In fact, the region to the left of point 6 is referred to as the *backside of the power curve* and is a region to be avoided, particularly on takeoff or landing. If the airspeed is allowed to drop below V_6 , it is not always possible to increase the thrust (by increasing the throttle setting) sufficiently or rapidly enough to prevent stalling, and increasing the thrust by losing altitude is not an acceptable alternative.

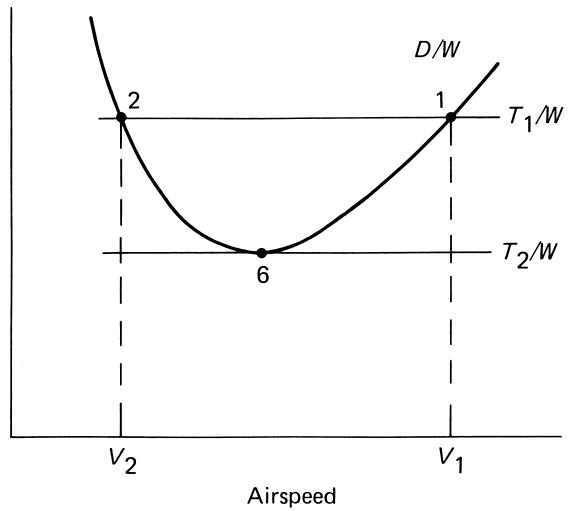


FIGURE 4 Level-flight equilibrium conditions for a turbojet for a given altitude and wing loading.

If the available thrust is reduced, the two airspeed solutions approach one another, coming together at the point where the available thrust curve is tangent to the drag curve, such as at point 6. This point represents the absolute ceiling of the aircraft for this throttle setting. If the thrust is further reduced, the available thrust will be less than the drag. The airspeed will begin to drop off, decreasing the lift, and the aircraft will lose altitude and descend until the available thrust again becomes equal to the drag. (With a constant throttle setting, the available thrust will increase as the altitude is lowered.)

In Eq. (19), T , the available thrust, is determined by the throttle setting and the altitude (and, of course, by the size of the engines). This equation shows that a high airspeed calls for a large T/S , where $T/S = (T/W)(W/S)$, and a small C_{D0} . With a maximum throttle setting and a constant weight, a theoretical flight envelope that shows the regions of possible level-flight airspeeds as a function of altitude can be constructed, as shown in Fig. 5. The dashed line represents the envelope using the conservative assumption that the thrust is directly proportional to the atmospheric density, whereas the solid line represents a more realistic (but less convenient) variation of the available thrust with altitude. The flight envelope should be further refined (and restricted) by superimposing lines for the stall speed and the drag-divergence Mach number as a function of the altitude.

If the thrust is the maximum available, then the absolute-ceiling condition for a given throttle setting [Eq. (20)] becomes the condition for the absolute ceiling of the aircraft and can be written:

$$T_{m,c}/W_c = 1/E_m \quad (21)$$

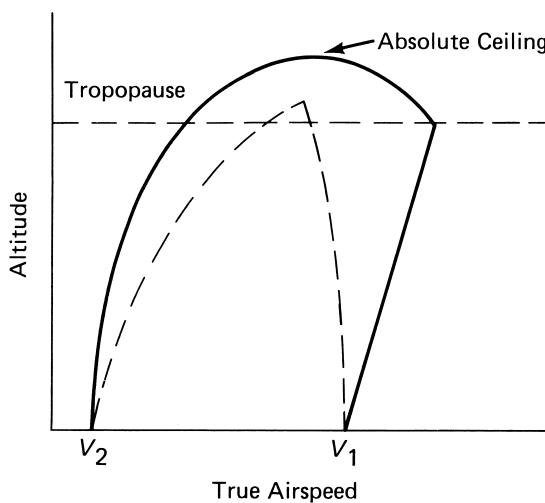


FIGURE 5 Level-flight airspeed-altitude flight envelope for a turbojet for a given weight and throttle setting.

With the maximum thrust at the ceiling approximated by the relationship $T_{m,c} = T_{m,SL}\sigma_c$, and with the assumption that the aircraft weights at sea level and at the ceiling are the same, the density ratio at the ceiling can be approximated by:

$$\sigma_c = \frac{1}{E_m(T_m/W)_{SL}} \quad (22)$$

which shows that the maximum lift-to-drag and thrust-to-weight ratios are the only design characteristics that affect the ceiling of a turbojet; the wing loading and the zero-lift drag coefficient, however, do influence the airspeed at the ceiling.

Let us look at the ceiling conditions for a turbojet/turbofan with a $W/S = 100 \text{ lb/ft}^2$, a sea-level $T_m/W = 0.25$, and a drag polar $C_D = 0.016 + 0.0468 C_L^2$, so that $E_m = 18.27$ and $M_{DR} = 0.82$.

For this aircraft, the ceiling density ratio is 0.219, which (using Table I) indicates an absolute ceiling of the order of 42,600 ft above mean sea level. This is a minimum ceiling because of the approximations and assumptions and because the ceiling will increase as fuel is consumed and the aircraft weight decreases, thus increasing the thrust-to-weight ratio. An aircraft with a high wing loading may not be able to reach its theoretical ceiling because the theoretical ceiling airspeed may exceed the drag-rise Mach number. An expression for the ceiling airspeed can be found from Eq. (19) by setting the radicand equal to zero. Notice the influence of the wing loading (in the T/S ratio) and of C_{D0} . The required airspeed at the ceiling for the aircraft is 831.5 fps or M 0.859, which exceeds the drag-rise Mach number and means that the actual ceiling will be lower than that calculated and the flight envelope will have a flattened top at the actual ceiling.

The service ceiling, which is defined as the altitude at which the maximum rate of climb is 100 fpm, is commonly used as a performance and design specification. The cruise ceiling is defined as the altitude at which the maximum rate of climb is 300 fpm. These ceilings are obviously lower than the absolute ceiling and are in the vicinity of the ceiling obtained from the approximation of Eq. (22).

Cruise range is the horizontal distance X traveled with respect to the surface of the earth and does not include any distances traveled during climb or descent. With the customary no-wind assumption, the cruise range is often called the still-air range. Returning to Eq. (15a), the T/W ratio can be shown to be equal to the D/W ratio, which in turn is equal to D/L or $1/E$, where E is the flight or instantaneous lift-to-drag ratio, not the maximum lift-to-drag ratio, and is a function of both the design characteristics and the flight conditions. The value of the flight lift-to-drag ratio is established by the lift coefficient required to maintain level flight at a specified airspeed and

altitude. Obviously, the minimum thrust required from the engines occurs when the aircraft is flying at the maximum value of the lift-to-drag ratio, and, just as obviously, the larger the maximum lift-to-drag ratio (the aerodynamic efficiency), the lower is the minimum required thrust.

In Eq. (15b), dividing dX/dt by dW/dt and making use of the various relationships yields the expressions:

$$\frac{dX}{-dW} = \frac{V}{cT} = \frac{V}{cD} = \frac{VE}{cW} \quad (23)$$

where $dX/-dW$ is often referred to as the instantaneous range or the specific range and is the exchange ratio between range and fuel. It has units of miles per pound of fuel and is analogous to the mileage of an automobile, especially when it is expressed in miles per gallon (mpg). A gallon of jet fuel weighs ~ 6.75 lb.

Although the mileage of an aircraft at a particular point or instant of time is of interest, the range for a given amount of fuel or, conversely, the amount of fuel required to fly a specified range is of greater interest in performance analyses. The range can be found by integrating the instantaneous performance over the interval between specified end points, usually the start and end of cruise, which are denoted by the subscripts 1 and 2, respectively. Before integrating Eq. (23), it is necessary to define the mass ratio MR and the cruise-fuel weight fraction ζ , both important performance and design parameters. The mass ratio is the ratio of the total weight of the aircraft at the start of cruise (W_1) to its total weight at the end of cruise (W_2) and is always greater than unity. The cruise-fuel weight fraction ζ is the ratio of the weight of fuel consumed during cruise (ΔW_f) to the total weight of the aircraft at the start of cruise; that is, $\zeta = \Delta W_f/W_1$. The relationships among the mass ratio, the cruise-fuel weight fraction, and the aircraft weights are

$$MR = \frac{W_1}{W_2} = \frac{W_1}{W_1 - W_f} = \frac{1}{1 - \zeta} \quad (24a)$$

$$W_2 = W_1(1 - \zeta) \quad (24b)$$

$$\zeta = (MR - 1)/MR \quad (24c)$$

With the assumption of a constant tsfc, Eq. (23) can be partially integrated to obtain the integral range equation:

$$X = \frac{-1}{c} \int_1^2 \frac{VE}{W} dW \quad (25)$$

Before this equation can be further integrated, it is necessary to define the flight program to be considered with two parameters held constant. Of the many flight programs that could be defined, one of particular interest is the constant airspeed-constant lift coefficient flight program. Since C_L is held constant, the lift-to-drag ratio E will also be constant, so that Eq. (25) can be written:

$$X = -\frac{VE}{c} \int_1^2 \frac{dW}{W} \quad (26)$$

The resulting range equation is

$$X = \frac{VE}{c} \ln MR = \frac{VE}{c} \ln \left(\frac{1}{1 - \zeta} \right) \quad (27)$$

This is the general form of what is known as the *Breguet range equation*. The airspeed and lift coefficient are related by the expression obtained from the lift equation, namely,

$$V = \left[\frac{2(W/S)}{\rho_{SL} \sigma C_L} \right]^{1/2} \quad (28)$$

In order to keep both V and C_L constant while the weight of the aircraft decreases along the flight path as fuel is used, Eq. (28) shows that the atmospheric density ratio must decrease in a similar manner so as to keep W/σ constant. The only way that this can be done is to increase the altitude (decrease σ in an appropriate manner). Consequently, the aircraft must maintain a continuous climb, and this flight program is commonly referred to as *cruise-climb* flight. Although this climb appears to violate the level-flight assumption, it will be shown that the flight-path angle is sufficiently small to justify the use of the level-flight equations and solutions for cruise-climb. The required thrust (the drag) will decrease along the flight path in such a manner that in the stratosphere the available thrust will decrease in an identical manner. Therefore, cruise-climb in the stratosphere requires no computations or effort on the part of the pilot. After establishing the desired cruise airspeed, the pilot simply engages the Mach-hold (or constant-airspeed) mode on the flight control system, and the aircraft will slowly climb at the flight-path angle required to keep the airspeed and lift coefficient constant.

Example 3. A turbofan with the drag polar of Example 1 is to cruise-climb at 455 mph (667 fps, M 0.67), starting at 30,000 ft with an initial wing loading of 130 lb/ft², a sfc = 0.65 hr⁻¹, and a cruise fuel fraction of 0.3 (a relatively long-range flight).

1. Find the range in miles:

$$q_1 = \frac{1}{2} \rho_{SL} \times 0.374 (667)^2 = 197.7 \text{ lb/ft}^2$$

$$C_{L1} = (W/S)/q = 130/197.7 = 0.658$$

$$C_{D1} = 0.016 + 0.0468 (0.658)^2 = 0.036$$

$$E_1 = C_L/C_D = 18.1$$

$$X = \frac{18.1 \times 455}{0.65} \ln \left(\frac{1}{1 - 0.3} \right) = 4519 \text{ mi} = 3924 \text{ nmi}$$

2. Find the altitude at the end of cruise and the average flight-path angle in degrees:

$$\sigma_2 = \sigma_2(1-\zeta) = 0.374 \times 0.7 = 0.262$$

From Table 1,

$$h_2 \cong 38,750 \text{ ft}; \quad \gamma = \tan^{-1} \frac{8750}{4519 \times 5280} = 0.021^\circ, \\ \text{a very small climb angle}$$

3. Find the flight time (hours):

$$t = 4519/455 = 9.93 \text{ hr}$$

It is important to watch the units. Since c is expressed in pounds per hour per pound and since the range is in statute miles, V in the range equation must be in miles per hour. If the range were to be in nautical miles (nmi), V would be expressed in knots, where 1 knot is 1 nmi/hr and 1 nmi = 1.1515 mi.

The range obtained in Example 3 is not necessarily the best range, that is, the maximum range possible for that starting altitude and fuel-flow weight fraction. The best-range conditions can be found by maximizing the instantaneous range (the mileage) of Eq. (23) by setting its first derivative with respect to the airspeed equal to zero and solving for the best-range airspeed. With a parabolic drag polar, the best-range airspeed is

$$V_{\text{br}} = \left[\frac{2(W/S)}{\rho_{\text{SL}} \sigma} \right]^{1/2} \left[\frac{3K}{C_{D0}} \right]^{1/4} \quad (29a)$$

and other best-range conditions are

$$C_{L,\text{br}} = \left(\frac{C_{D0}}{3K} \right)^{1/2}; \quad E_{\text{br}} = 0.866E_m \\ \frac{T_{\text{br}}}{W} = \frac{1.155}{E_m} \quad (29b)$$

Example 4. As in Example 3, the turbofan flies a cruise-climb program starting at 30,000 ft with $W/S = 130 \text{ lb/ft}^2$ and $\zeta = 0.3$ but at “best-range” conditions.

1. Using the best-range expressions above, $V_{\text{br}} = 930.8 \text{ fps} = 634.6 \text{ mph} = M 0.93 > M_{\text{DR}}$ (0.83), and drag polar is *not* valid. Since $M_{\text{br}} > M_{\text{DR}}$, cruise at $M_{\text{DR}} = 0.83$; therefore, $V = 825 \text{ fps} = 562.7 \text{ mph}$; $q = 302.8 \text{ lb/ft}^2$; $C_L = 130/302.8 = 0.429$; $C_D = 0.0246$; and $E = 17.4$.

2. Find the range and flight time:

$$X = \frac{562.7 \times 17.4}{0.65} \ln \left(\frac{1}{1 - 0.3} \right) = 5372 \text{ mi} \\ = 4666 \text{ nmi} \text{ (a 19% increase in range)} \\ t = 5372/562.7 = 9.54 \text{ hr} \\ \text{(a 4% decrease in flight time)}$$

3. As a matter of interest, find the cruise-fuel weight fraction for a range of 2000 mi.

$$2000 = \frac{562.7 \times 17.4}{0.65} \ln MR; \quad MR = 1.142; \\ \zeta = \frac{MR - 1}{MR} = 0.124$$

First, notice that it is not possible to fly at the best-range conditions because the large wing loading of this aircraft requires a best-range airspeed that exceeds the drag-rise Mach number and violates the parabolic drag polar assumption; however, this example shows that the benefits of flying at the higher airspeed with the same fuel load are an increase in range of 19% and a 4% reduction in flight time. Also notice the coupling of the airspeed and the lift-to-drag ratio in the range equation; one cannot be changed without affecting the other, and the range for a given fuel load is maximized when the product of V and E is maximized.

Although V and E are coupled for a given σ and W/S , if E is specified, such as E_{br} , Eq. (29a) shows that V_{br} can be increased without changing E by decreasing σ (by increasing the altitude) and/or increasing W/S . The influence of the altitude on the airspeed and range of a turbojet/turbofan is quite strong, as evidenced by the fact that the best-range airspeed (and best range) for a particular aircraft is 60% higher at 30,000 ft than at sea level. Although the effects of increasing the wing loading are not as dramatic, a larger W/S increases both the airspeed and range; for example, a doubling of the wing loading results in a 14% increase in both the airspeed and the range.

Typical values of the wing loading are of the order of 100–130 lb/ft² for long-range subsonic transports, of the order of 50 to 60 lb/ft² for short- and medium-range transports and fighter aircraft, and of the order of 15–20 lb/ft² for light aircraft. The lower values for the last two classes of aircraft arise from operational requirements more pressing than range, generally a short takeoff run for the smaller aircraft and maneuverability for the fighters.

Two cautionary comments are in order with respect to increasing the best-range airspeed. First, no attempt should be made to increase the wing loading by increasing the weight; instead, the wing area should be reduced. Although V_{br} is directly proportional to the square root of the gross weight, the instantaneous range is inversely proportional to the weight itself. Second, no attempt should be made to increase V_{br} by increasing K (by decreasing the aspect ratio), because the range penalty associated with the decrease in E_m will be predominant. On the other hand, reducing the zero-lift coefficient will increase not only V_{br} but also E_m , both salutary. Furthermore, a reduction in C_{D0} improves all other aspects of performance and is a goal to be pursued vigorously, as is a reduction in weight.

Equation (29b) shows that the best-range T/W is greater than $1/E_m$, as is required for level flight, but not much more, and it can be shown that the cruise-climb altitude is ~ 3500 ft lower than the absolute ceiling for the best-range throttle setting. Equation (29b) also shows that the thrust required is independent of the altitude itself so that the sea-level thrust (as is the drag) is identical to that required at high altitudes. As the altitude increases, however, the available thrust decreases, and it is necessary to be sure that there is sufficient available thrust at the cruise altitude and that the best-range airspeed does not exceed the drag-rise Mach number. It can be easily shown that the best-range fuel consumption is also independent of the altitude. This means that, for a given fuel load, a turbojet aircraft can stay in the air just as long at sea level as at altitude. It is the increased airspeed required at the higher altitudes that results in the greater ranges. In actuality, the tsfc does decrease slowly with altitude, reaching a minimum value at the tropopause and increasing even more slowly in the stratosphere. Consequently, there is a slight advantage to flying in the vicinity of the tropopause, all other things being equal.

Although the cruise-climb flight program is superior to any other cruise flight program, air traffic control will permit its use only under certain limited conditions because the continuous climb poses a threat to other traffic. For short-range flights, flying at a constant altitude with a constant airspeed, which is favored by air traffic control, does not significantly penalize the range performance. On long-range flights, cruise-climb can be approximated by the use of stepped-altitude flight, which is a series of constant altitude-constant airspeed segments conducted at increasing altitudes with each altitude change placing the aircraft momentarily on the cruise-climb flight path. Stepped-altitude flight, with 4000-ft steps (the smallest allowed by air traffic control), is consistently used on transoceanic and some transcontinental flights to reduce the total fuel consumption, an important consideration in times of fuel scarcity and/or increasing prices.

Endurance is the length of time an aircraft can remain airborne for a given expenditure of fuel and for a specified set of conditions. The instantaneous endurance, or exchange ratio for time and fuel, is simply the inverse of the negative fuel consumption rate and, for level or cruise-climb flight, is simply $dt/-dW$, which is equal to E/cW . Maximum endurance is a primary design and operational criterion for aircraft with such special missions as patrol, antisubmarine warfare, and command and control and is of interest for all aircraft during the loiter phase of any flight. Loiter is defined as flight where endurance is paramount and range is either secondary or of no importance at all. An intercontinental bomber, such as the B-52, loiters while on airborne alert. A fighter on com-

bat air patrol (CAP) loiters while awaiting the assignment of targets or the sighting of intruders. All aircraft should loiter, if so permitted, in holding patterns while awaiting further clearance.

Inspection of the instantaneous endurance shows that it is at maximum when the weight is at a minimum and the aircraft is flying at E_m , that is, when the drag is at a minimum and the T/W is equal to $1/E_m$. The flight altitude will be the absolute ceiling for the throttle setting and the airspeed will be 24% lower than that for best-range cruise at the same altitude. The maximum endurance is independent of the altitude, but the associated range will be higher at altitude because the minimum-drag airspeed will be higher at altitude. It is interesting that the maximum-endurance flight time is always 15.5% longer than the best-range flight time, regardless of the magnitude of the range.

B. Other Flight in the Vertical Plane

Takeoff and landing ground runs, though not flight conditions per se, are important for performance and operational reasons and often are the determining factors in the selection of certain design and subsystem characteristics, such as the thrust-to-weight ratio and the wing loading.

The equation of motion governing the takeoff ground run of a conventional aircraft can be written

$$T - D - \mu(W - L) = \frac{W dV}{g dt} \quad (30)$$

where μ is the coefficient of rolling friction and g the acceleration due to gravity. Although numerical solutions or iterative procedures are required for the solution of this equation, a reasonable closed-form solution that gives some insight to the takeoff problem can be obtained by assuming that the thrust is constant throughout the run and by neglecting the drag and friction forces, which account for 10–20% of the energy expended. With these assumptions, Eq. (30) can be simplified and written

$$T = \frac{W}{g} \frac{dV}{dx} = \frac{WV}{g} \frac{dV}{dx} \quad (31)$$

Separating the variables and integrating from the start of the run to liftoff (LO) yields the expression:

$$d = \frac{V_{LO}^2}{2g(T/W)} \quad (32)$$

The liftoff airspeed V_{LO} is generally set equal to $1.2 V_s$, where V_s is the takeoff stall speed. Expressions for the two airspeeds are

$$V_s = \left[\frac{2(W/S)}{\rho_{SL} \sigma C_{Lmax, TO}} \right]^{1/2} \quad (33a)$$

$$V_{LO} = 1.2V_s = \left[\frac{2.88W/S}{\rho_{sc}\sigma C_{Lmax,TO}} \right]^{1/2} \quad (33b)$$

Examination of Eqs. (32) and (33b) shows that the takeoff run can be decreased by decreasing the wing loading or by increasing the thrust-to-weight ratio, the maximum-lift coefficient, or the atmospheric density. Decreasing the wing loading, however, decreases the best-range airspeed and thus the cruise range for a given cruise-fuel weight fraction. Increasing the thrust-to-weight ratio beyond that required for cruise or for special mission requirements (such as ceilings) increases the engine weight with accompanying increases in the gross weight of the aircraft and fuel required. Increasing the maximum-lift coefficient by means of flaps (Fowler-type flaps also decrease the wing loading by increasing the wing area) adds weight and complexity and increases the drag during the ground run. The atmospheric density cannot be controlled, but its effect on takeoff performance cannot be neglected. Obviously, an airfield located 5000 ft above mean sea level will require a longer ground run than one at sea level. In addition, a hot day and/or low barometric pressure will also decrease the air density, thus increasing the ground run by increasing the liftoff airspeed and decreasing the available thrust.

The thrust-to-weight ratio, although appearing to be dimensionless, is actually an acceleration expressed in g 's. In other words, a T/W of 0.25 represents an acceleration of 0.25 g , or 8.05 ft/sec². Therefore, it is possible to develop an approximate, and somewhat optimistic, expression for the time to liftoff, that is,

$$t = \frac{V_{LO}}{g\sigma(T/W)_{SL}} \quad (34)$$

A long-range, high-subsonic transport with a wing loading of 100 lb/ft², a T/W of 0.25, and a maximum-takeoff-lift coefficient of 1.8 would require (according to the preceding equations) a sea-level ground run of 4200 ft with a liftoff airspeed of 177 mph and an elapsed time of 32 sec. Given that these are optimistic figures and that provisions must be made for nonstandard day conditions, clearance of a 50-ft obstacle, factors of safety, aborts, and so on, such a ground-run value implies the need for runway lengths of the order of 8000–10,000 ft for such aircraft.

Short-range transports used for travel between and from smaller cities cannot count on finding 10,000-ft runways. Consequently, their design characteristics must be modified so as to reduce the ground run, generally by reducing the wing loading and increasing the lift coefficient. The T/W is left unchanged, if possible, for weight and operating economy reasons. Reducing W/S from 100–50 lb/ft² and increasing the lift coefficient from 1.8–2.0, but leaving the T/W at 0.25, results in a ground run of ~1900 ft with a liftoff airspeed of 119 mph and an elapsed time of

22 sec. This ground-run value implies a need for runway lengths of the order of 4000 to 5000 ft.

A STOL (short takeoff and landing) aircraft is designed for exceptionally short ground runs as well as steep climbs and descents. It is characterized by a low wing loading, a high lift coefficient, and a large thrust-to-weight ratio. For example, if W/S is 40 lb/ft², C_L is 2.6, and T/W is 0.6, then the ground run becomes of the order of 500 ft with a liftoff airspeed of 93 mph and an elapsed time of 7 sec.

Air-to-air fighter aircraft, particularly those equipped with guns, are characterized by a lower wing loading than transports (for reasons of maneuverability), higher T/W values, as might be expected, and lower values of the maximum-lift coefficient (to keep the wing thin and clean so as to minimize C_{D0} and the wave drag). An advanced fighter with a T/W of 1.3 (without afterburner), a W/S of 66 ft/lb², and a maximum C_L of 1.2 would have a ground run of the order of 800 ft and an elapsed time of 6 sec. The sea-level stall and liftoff airspeeds would be 215 fps (147 mph) and 258 fps (176 mph), respectively.

The landing-run requirements must also be considered. They have increased in importance as wing loadings (and thus approach and touch-down airspeeds) have increased and as the drag has decreased. The landing maneuver comprises the final approach, the landing flare, the touch-down (to include getting all the wheels on the ground), and the ground run. A simple but crude expression for the landing-run distance, in feet, is

$$d \cong \frac{1}{2}f_d V_s^2 \quad (35)$$

where f_d , the deceleration factor, is of the order of 0.4 for a conventional jet transport and V_s is the stall speed at touchdown. This expression is crude because of the variety and complexity of modern retardation devices (such as thrust reversers and spoilers) and the difference between wet and dry runways. If the illustrative turbojet has been on a long-range mission with a cruise-fuel weight fraction of 0.3 and has a landing lift coefficient of 2.0, then the landing wing loading will be equal to 100(1 – 0.3), or 70 lb/ft²; V_s will be 171.6 fps (117 mph) with a final-approach air-speed of 140 mph. With an f_d of 0.4, the landing run will be of the order of 5890 ft, which is longer than the takeoff run. Stopping a clean, high-speed aircraft is not a trivial task. In fact, the landing run often establishes the minimum final-approach airspeed, which in turn fixes the maximum value of the wing loading.

The climbing flight steady-state equations are

$$T - D - W \sin \gamma = 0 \quad (36)$$

$$L - W \cos \gamma = 0 \quad (37)$$

$$dX/dt = V \cos \gamma \quad (38)$$

$$dh/dt = V \sin \gamma \quad (39)$$

where γ , the climb (flight-path) angle is the angle between the horizon and the velocity vector. From Eq. (36):

$$\sin \gamma = (T - D)/W \quad (40)$$

which shows that the climb angle is determined by the excess thrust per unit weight, where *excess thrust* refers to that thrust not required to counteract the drag. Substituting Eq. (40) into Eq.(39) gives an expression for dh/dt , the rate of climb R/C . This expression.

$$R/C = V \sin \gamma = (TV - DV)/W \quad (41)$$

shows that the rate of climb is determined by the excess thrust power per unit weight.

Rewriting Eq. (40) and incorporating Eq. (37) yields:

$$\sin \gamma = \frac{T}{W} - \frac{D}{W} = \frac{T}{W} - \frac{\cos \gamma}{E} \quad (42)$$

which shows that the largest value that the sine of the climb angle can assume is equal to the maximum thrust-to-weight ratio. If the maximum T/W is 0.25, a typical value for a subsonic jet transprot, then the maximum climb angle will be less than 15° at sea level; if the maximum T/W is 0.6, the maximum climb angle will be less than 37° . As the altitude increases during the climb, the available T/W will decrease (with W changing slowly) and the maximum climb angle will therefore decrease, going to zero at the absolute ceiling of the aircraft, as might be expected.

With respect to the climb drag, it can be shown that, for climb angles up to the order of 30° , using the level-flight drag is a reasonable approximation. The fuel burned during a climb is a small percentage of the gross weight of a well-designed aircraft climbing in an efficient manner, and the aircraft weight will be assumed constant.

Steepest climb is of interest in clearing obstacles and in establishing the upper limit of the climb angles and occurs when the excess thrust is at a maximum. With the level-flight drag approximation, steepest climb conditions are essentially the minimum drag (E_m) conditions as can be seen in Fig. 6. For the illustrative turbojet at sea level, the maximum climb angle is 11.3° at an airspeed of 259 mph with an R/C of 4460 fpm. At 30,000 ft, the maximum climb angle has dropped to 2.2° , the airspeed has increased to 423 mph, and the R/C is 1443 fpm.

Fastest climb is of greater importance as it requires the minimum time to climb to a specified altitude, which is of interest to air traffic controllers, who must keep the air space free of other traffic, and to a first approximation requires the smallest amount of fuel, thus increasing the amount available for cruise. The R/C can be maximized by setting its first derivative with respect to the true airspeed equal to zero, yielding the following relationships:

$$V_{FC} = \left[\frac{(T/S)_{SL} \Gamma}{3 \rho_{SL} C_{D0}} \right]^{1/2} \quad (43)$$

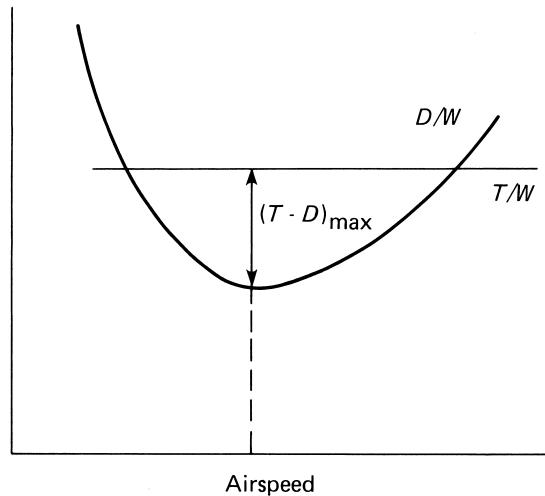


FIGURE 6 Graphic representation of steepest climb condition for a turbojet for a given wing loading, altitude, and throttle setting.

$$\Gamma = 1 + \left\{ 1 + \frac{3}{[E_m(T/W)]^2} \right\}^{1/2} \quad (44)$$

$$\sin \gamma_{FC} = \frac{T}{W} \left(1 - \frac{\Gamma}{6} \right) - \frac{3}{2\Gamma E_m^2(T/W)} \quad (45)$$

$$(R/C)_{max} = V_{FC} \sin \gamma_{FC} \quad (46)$$

A close examination of these expressions shows that a high rate of climb calls for a large T/W , a high W/S , and a low C_{D0} and that the magnitude of the thrust is the most important parameter for fast climb. Fastest climb of the illustrative turbojet at sea level is 6266 fpm at 459 mph (M 0.6) with a climb angle of 8.9° . This airspeed is faster than that allowed by air traffic control at the lower altitudes and would have to be reduced appropriately.

It should be emphasized that these climbing performance values are theoretical and represent the maximum performance possible without any practical limitations on the climbing airspeed, as mentioned in the previous sentence, or on the maximum allowable load factor, to be discussed in the next section on turning flight.

The time to climb and the amount of fuel used during climb are of more interest than the R/C and can be found from approximate closed-form expressions or by dividing the climb into segments and then using average values to create a tabular solution.

Unpowered flight occurs when a single-engine aircraft has engine failure, when an aircraft runs out of fuel, or when an aircraft has no engine (a glider or a sailplane). The relevant equations of motion are obtained by setting the thrust equal to zero and assuming a small glide (flight-path) angle. Maximum range occurs when the aircraft is flown at its maximum lift-to-drag ratio (which is often

referred to as the glide ratio). Since the glide angle is constant and has its smallest value, this glide program is sometimes called flattest glide. The airspeed will be the minimum drag airspeed and is higher than one might expect. Maximum-endurance (minimum rate of descent) glide calls for a 15% larger glide angle, a slower airspeed, and an E equal to $0.866E_m$.

C. Turning Flight

Even though an aircraft may spend most of a mission in straight flight, there are times when it must turn. For all aircraft there are turns associated with changes in flight headings, collision avoidance, and instrument approaches and landings. Combat aircraft must have a greater degree of maneuverability than transports in order to survive and carry out their operational missions. The maneuvering capability of an aircraft, along with the associated design and flight parameters, can be determined to a first approximation by limiting turning flight to the horizontal plane and by assuming the weight to be constant and the sideslip angle to be zero. The maneuvering parameters of interest are the (maximum) turning rate and the (minimum) turning radius.

The relevant steady-state equations of motion are

$$T = D \quad (47)$$

$$L \sin \phi = \frac{W}{g} V \dot{\chi} \quad (48)$$

$$L \cos \phi = W \quad (49)$$

where $\dot{\chi}$ is the turning rate (rad/sec) and ϕ the bank angle (degrees). These equations show that the drag is balanced by the thrust, the centrifugal force is balanced by the horizontal component of the lift, and the weight is balanced by the vertical component of the lift. With the sideslip angle zero (coordinated turns), the lift vector always lies in the plane of symmetry.

The lift-to-weight ratio L/W is called the load factor is given the symbol n , and has units of g 's. It is called the load factor because it is a measure of the forces, or loading, impressed on the structure or occupants. For example, level flight is often referred to as $l-g$ flight because $L = W$ and $n = 1$. When n is equal to 2, the lift is twice the weight of the aircraft and the wing span, for example, must accept a load that is twice the weight of the aircraft without damage or unacceptable deflections. Similarly, a person in the aircraft is subjected to an additional force equal to the individual's weight. If the load factor exceeds the tolerances of the structures or occupants, temporary or permanent damage can occur.

Using the definition of the load factor and Eq. (47), one can write Eq. (49) as:

$$n = L/W = 1/(\cos \phi) = (T/W)E \quad (50)$$

which shows a direct coupling with the bank angle and with the product of the thrust-to-weight and lift-to-drag ratios. For any load factor, the corresponding bank angle is

$$\phi = \cos^{-1}(1/n) \quad (51)$$

When n is unity, the lift is equal to the weight and the bank angle is zero. When the bank angle is zero, Eq. (48) shows that the turning rate is also zero, indicating that there cannot be a steady-state coordinated turn without a finite bank angle. When the bank angle is 90° , the load factor becomes infinite, which means that an aircraft cannot maintain a coordinated steady-state turn with a 90° bank angle and still hold a constant altitude. Turns with a bank angle of 90° can be accomplished only by skidding or by losing or gaining altitude, the latter requiring a T/W greater than 1.

Relevant expressions for $\dot{\chi}$ are

$$\dot{\chi} = \frac{g \tan \phi}{V} = \frac{g(n^2 - 1)^{1/2}}{V} \quad (52)$$

where $\dot{\chi}$ is in radians per second, V is in feet per second, and g is taken to be equal to 32.2 ft/sec^2 . Since the airspeed V is a tangential velocity equal to $\dot{\chi}r$, expressions for r are

$$r = \frac{V}{\dot{\chi}} = \frac{V^2}{g \tan \phi} = \frac{V^2}{g(n^2 - 1)^{1/2}} \quad (53)$$

The D/W ratio, using the parabolic drag polar, can be written:

$$\frac{D}{W} = \frac{q C_{D0}}{W/S} + \frac{Kn^2(W/S)}{q}. \quad (54)$$

which, when set equal to T/W , can be solved for the turning airspeed,

$$y = \left\{ \frac{T/S}{\rho_{SL}\sigma C_{D0}} \left[1 \pm \left(1 - \frac{n^2}{[E_m(T/W)]^2} \right)^{1/2} \right] \right\}^{1/2} \quad (55)$$

These expressions have several points of interest. When the lift is equal to the weight, the load factor is unity, and the turning values reduce to those for level flight, as might be expected. When the load factor is greater than unity, as in a turn, the airspeed will be less than that for level flight. Consequently, the pilot must increase the thrust when entering a turn if both the airspeed and altitude are to remain constant. It should also be noted that specifying the bank angle (and thus the load factor) and the T/W ratio determines the airspeed and turning rate for a given altitude and that there may be two possible combinations.

Not only does the airspeed drop off in a turn if the thrust is not increased, but also the stall speed increases as the square root of the load factor. For example, a bank angle of 30° ($n = 1.155$) increases the wing-level stall speed by 7.5%.

Solving the turning rate and turning radius equations for several airspeed combinations leads to the conclusion that *lower airspeeds and lower altitudes give better turning performance* in terms of higher turning rates and smaller turning radii.

Inspection of Eq. (50) shows that for a given T/W , the maximum load factor (and maximum bank angle) occurs when E is at a maximum and that the largest value of the maximum possible load factor occurs when both T/W and E are at their maxima. Because of structural limitations (or passenger comfort), aircraft are limited to flight conditions that will not exceed a maximum allowable load factor, which for a transport is of the order of 2.5 g and for a fighter aircraft is of the order of 8 g . The illustrative turbojet has a maximum possible load factor at sea level of 4.6 g . If the aircraft is not to exceed 2.5 g , the T/W ratio must not exceed 0.137 g . The maximum possible load factor decreases with altitude (decreasing T/W), becoming unity at the absolute ceiling, so that the maximum bank angle is zero. An aircraft cannot make a steady turn at the absolute ceiling without losing altitude.

Maximum turning rate (fastest turn) and minimum turning radius (tightest turn) are the true measures of the maneuverability of an aircraft. Both are achieved at low airspeeds and at sea level. As a consequence, air-to-air combat may start at M 2 to M 3 and at 50,000–60,000 ft but, if continued, will slow down and descend until the low speed–low altitude combat arena is reached or until one of the aircraft is destroyed. The lower limit to the combat airspeed is the stall speed in the turn, and the lower limit to the combat altitude is the minimum altitude required for pull-up from an evasive maneuver. A highly maneuverable aircraft should have a large T/W , a low W/S , a large aspect ratio and Oswald span efficiency, a low C_{D0} , and a large maximum-lift coefficient. Most of these characteristics can be seen in a modern aircraft designed for maneuvering combat with the exception of the large aspect ratio, which is kept small for structural and weight considerations.

There are two differences between fastest turns and tightest turns. The fastest turn load factor is proportional to the square root of the maximum possible load factor and has no theoretical upper limit, whereas the tightest turn load factor cannot exceed 1.414 g . The tightest turn airspeeds are even lower than those for fastest turn and usually less than the stall speed for fighter aircraft, a physically impossible situation.

V. PISTON-PROPELLER AIRCRAFT

Although the techniques used to evaluate the performance of turbojet aircraft can be applied directly to aircraft equipped with reciprocating engines and propellers (piston–props), their application is not as straightforward or as simple. Since space is too limited for complete developments and analyses of the various aspects of piston–prop performance, the approach used in this section is essentially one of comparison, with emphasis on the differences between the two classes of aircraft. The flight conditions for best performance and the design parameters are surprisingly different for the piston–prop.

The differences between the two arise from the fact that, whereas the turbojet engine produces thrust only, the piston–prop produces power only. (The analyses and comparative comments of this section apply to any conceivable propulsion system that delivers power only.) The difficulties in applying the turbojet techniques to the piston–prop analyses have two root causes. The first is that the fuel consumption rate of a piston–prop aircraft is proportional to the brake horsepower of the engine. The second, and more troublesome mathematically, is the fact that it is the thrust rather than the power that appears explicitly in the force equations.

For level flight in the vertical plane (the flight-path angle is zero) the equations of interest can be written:

$$P = TV = DV; \quad L = W \quad dX/dt = V \quad (56a)$$

$$\frac{dW}{dt} = -\hat{c}HP = \frac{-\hat{c}P}{k\eta_p} = \frac{-\hat{c}TV}{k\eta_p} \quad (56b)$$

In Eq. (56a), P and TV represent the available thrust power, and DV represents the required (drag) power. Useful power-to-weight ratio expressions are

$$P/W = DV/L = V/E \quad (57)$$

where E is the flight (instantaneous) lift-to-drag ratio.

Introducing the parabolic drag polar into Eq. (57) yields an expression that establishes the level-flight relationships between the thrust power and the airspeed in terms of the aircraft parameters and altitude, namely,

$$\frac{P}{W} = \frac{\rho_{SL}\sigma V^3}{2(W/S)} + \frac{2K(W/S)}{\rho_{SL}\sigma V} \quad (58)$$

Unfortunately, there is no closed-form solution, and values of the airspeed must be obtained by iteration. The first term on the right-hand side of Eq. (58) shows that the zero-lift drag power (required power) increases as the cube of the airspeed, whereas the zero-lift drag (required thrust) of the turbojet increases as the square of the airspeed. This

cubing effect is an important factor in limiting the maximum airspeed of a piston-prop. For example, the maximum sea-level airspeed of a high-performance piston-prop with a HP/W ratio of 0.1 hp/lb (two 375-hp engines) is 249 mph. If the maximum airspeed is doubled, the new HP/W ratio is 0.731 hp/lb, an increase by a factor of 7.31, and 15 engines would be required (not very likely).

The absolute ceiling occurs when the available power is at a maximum and the required power is at a minimum, leading to an expression for the ceiling density ratio,

$$\sigma_c = \left[\frac{1.155 V_{P\min,SL} \sigma_{cr}}{(P_m/W)_{SL} E_m} \right]^{2/3} \quad (59a)$$

where

$$V_{P\min,SL} = \left[\frac{2(W/S)}{\rho_{SL}} \right]^{1/2} \left(\frac{K}{3C_{D0}} \right)^{1/4} \quad (59b)$$

In Eq. (59a), σ_{cr} is the critical density ratio for a turbocharged engine and should be set equal to unity for an aspirated engine. The ceiling of a piston-prop is dependent not only on the engine power-to-weight ratio and E_m (as was the case with the turbojet) but also on the wing loading. As the wing loading is increased to increase the cruise airspeed, the ceiling is lowered, whereas the ceiling of a turbojet is independent of the wing loading.

Instantaneous (point) range expressions can be written:

$$\frac{dX}{-dW} = \frac{k\eta_p V}{\hat{c}P} = \frac{k\eta_p E}{\hat{c}W} \quad (60)$$

Notice that this piston-prop mileage is explicitly independent of the airspeed, although the value of E is directly coupled to the airspeed. For good mileage, the propeller efficiency and lift-to-drag ratio should be as large as possible, and the aircraft weight and specific fuel consumption should be minimized. If the units of the mileage are miles per pound, k has a numerical value of 375; for miles per gallon, one multiplies by 6, the weight of 1 gal of gasoline. Although the propeller efficiency varies with the airspeed, with a well-designed, variable-pitch, constant-speed propeller, it can be assumed to be constant in the design operating airspeed range and to have a value between 80 and 85%.

With the general assumptions of constant propeller efficiency and specific fuel consumption along with a flight condition of constant lift coefficient, integration of the mileage expression yields the piston-prop version of the Breguet range equation,

$$X = \frac{375\eta_p E}{\hat{c}} \ln \left(\frac{1}{1-\zeta} \right) \quad (61)$$

This equation is valid for both constant-altitude and cruise-climb flight, but there is no improvement in range for cruise-climb of a piston-prop as is the case with a

turbojet. It is customary to use the Breguet range equation for the normal constant altitude–constant airspeed flight program.

By inspection, the best-range condition for a piston-prop is to cruise at the maximum lift-to-drag ratio (and at the ceiling for the required throttle setting). Note that the best range is independent of both the airspeed and the altitude but that the aircraft must fly at the minimum-drag (best-range) airspeed. Increasing the best-range airspeed by increasing the altitude and/or wing loading will not increase the range but will certainly reduce the flight time and increase the power required.

For piston-prop aircraft characterized by low gross weights and moderate or low wing loadings, the best-range airspeed is considerably lower than the maximum airspeed. Furthermore, the distances normally flown are of the order of 1000 mi or even less, so that the design cruise-fuel weight fraction is low. Consequently, it is customary to cruise at 75% of the maximum power available because time-of-flight savings are more important than any fuel savings.

The differences between the best-range performance of piston-prop and turbojet aircraft can be seen by comparing the respective Breguet range equations. An interesting and significant observation is that the $375\eta_p$ in the piston-prop equation is an equivalent best-range airspeed that is independent of the actual airspeed at which the aircraft is flying. This means that if two piston-props have identical E_m 's and specific fuel consumptions, a 100-mph aircraft will fly just as far as a 600-mph aircraft if the cruise-fuel weight fractions are also identical. The faster aircraft, obviously, will require only one-sixth of the time to make the flight.

This equivalent best-range airspeed also means that a pure turbojet with its higher specific fuel consumption is not competitive with the piston-prop at lower airspeeds. For comparable aircraft the best-range airspeed of the turbojet will be 31% higher than that of the piston-prop. The best range, however, will always be less than that of the piston-prop until the best-range airspeed satisfies a condition that is a function of the specific fuel consumptions, the E_m 's, and the propeller efficiency. For example, if a piston-prop has a propeller efficiency of 85%, a specific fuel consumption of 0.5 lb/hr·hp, and an E_m of 14 and the turbojet has a specific fuel consumption of 0.8 lb/hr·lb and an E_m of 18, the turbojet must have a wing loading-altitude combination such that the best-range airspeed will be 457 mph.

The condition for piston-prop maximum endurance is to fly at minimum power at the lowest possible airspeed. This implies an aircraft that has a low wing loading, a large maximum lift-to-drag ratio, and a low specific fuel consumption and that flies as low and as slowly as possible.

Since the maximum-endurance airspeed is at the start of the backside of the power curve, the operational airspeed is somewhat higher than the theoretical value so as to establish the cruise airspeed at a stable equilibrium point.

Comparing performance other than range is much less straightforward and satisfying because it is difficult to establish a true basis of comparison. Perhaps the best that can be done is to specify identical wing loadings and drag polars (although identical C_{D0} 's are very unlikely and modify any analytical comparisons with observations and comments on the pertinent characteristics of the two types of propulsion systems.

With this definition of comparable aircraft, the one with the larger T/W ratio should have the shorter takeoff ground run. Evaluating the respective T/W 's is not simple, particularly since the HP/W ratio and not the T/W ratio is the propulsion characteristic of the piston-prop. Typical values for transport aircraft are 0.25 for the T/W of a turbojet and 0.1 for the HP/W of a piston-prop. If the rule of thumb relationship that 1 hp produces ~ 2.5 lb of static thrust is applied to the piston-prop, the two T/W ratios will be the same and so, therefore, should be the takeoff runs. However, since the thrust of a turbojet falls off at low airspeeds, its actual takeoff T/W ratio will be smaller. If it is 80% of the nominal T/W ratio, then the takeoff run of the piston-prop will be of the order of 80% of that for a turbojet.

There are other complications to be considered, however. The first of these is the fact that, as the wing loading of the piston-prop is increased in order to obtain higher cruise speeds, the liftoff airspeed will increase, thus reducing the effective T/W ratio of the piston-prop. Another, and probably more important, consideration is that the T/W ratio of the turbojet can be increased with a significantly lower increase in weight than that associated with increasing the HP/W ratio of the piston-prop. After all of this discussion, the only conclusion that can be reached is that it is not possible to make a flat statement as to the takeoff superiority of either type of aircraft.

With respect to the maximum rate of climb, the turbojet has a decided superiority at higher wing loadings (of the order of $50 \text{ lb}/\text{ft}^2$ or more) and has higher climb airspeeds. At lower wing loadings or at the lower climb speeds imposed by air traffic control, the climb rates of the two aircraft are much closer together. In addition, the turbocharged piston-prop has the definite superiority of minimizing the effects of an increasing altitude on the climb rate.

Whereas the absolute ceiling of a turbojet is independent of the wing loading (as long as the ceiling airspeed does not exceed M_{DR}), that of the piston-prop decreases with an increase in the wing loading. For both types of aircraft, increases in E_m and in T/W or HP/W raise the

ceiling, with somewhat larger proportionate increases for the piston-prop. In general, the turbojet has a ceiling that is considerably higher than that of the aspirated piston-prop, but the turbocharged piston-prop is much more competitive. In fact, with a low wing loading, the turbocharged ceiling might even be higher.

Turning performance is described in terms of the maximum load factor, the fastest turn, and the tightest turn. Since the characteristics that determine the ceiling also determine n_{max} , the n_{max} of the turbojet will generally be higher than that of the piston-prop. Turbocharging, however, will not increase n_{max} ; it will merely retard the decrease with altitude. Looking at the fastest and tightest turns for comparable aircraft without regard for whether the aircraft is below the stall speed, the piston-prop seems to have a slight edge, an edge that diminishes with increased wing loading. The maximum achievable lift coefficient is always a key factor in determining maneuverability, regardless of the type of propulsion system.

It is difficult to compare these two classes of aircraft because of the fundamental differences in their characteristics. The piston-prop has better mileage than the turbojet but is restricted to the low-speed regime by the excessive weight of the large engines required for the higher airspeeds and by the decrease in the propeller efficiencies at high speeds. Furthermore, the complexity and maintenance problems and costs associated with large internal combustion engines are very high, and, finally, the increase in the wing loading required for the higher airspeeds adversely affects other aspects of performance.

Although the range of the piston-prop is in different to altitude and airspeed, the faster (and thus the higher) the turbojet flies, the farther it can fly with a given fuel load. The upper limit to subsonic cruise airspeed is the drag-rise Mach number, at which point E_m begins to fall off. The primary disadvantage of the turbojet is its large specific fuel consumption, and a secondary disadvantage is low thrust at very low airspeeds.

VI. TURBOPROP AND TURBOFAN AIRCRAFT

Turboprops and turbofans are gas turbine engines, as is the turbojet, and are designed to minimize the disadvantages of and exploit the advantages inherent in piston-prop and turbojet engines. The fundamental difference among these three engines is in how they produce thrust. The turbojet does it by expansion of hot gases through a nozzle, the turboprop uses a propeller, and the turbofan uses a multi-bladed fan, which is related in many ways to the propeller. The basic element of a gas turbine engine is the gas generator, made up of the compressor(s), the burners, and the

turbines that drive the compressor. The mixture of air and fuel that passes through the gas generator is the primary flow. The gas generator and primary flow are common to all three engines and serve as a baseline for a comparative evaluation.

In a turbojet, the exhaust gases from the gas generator expand through a nozzle and thrust is the only output. It is a single-flow engine whose distinguishing characteristics are its light weight, small frontal area, propulsive efficiency that increases with airspeed, high specific fuel consumption (the highest of the three), and low thrust at low airspeeds.

In a turboprop, there are two flows: a primary flow that develops jet thrust and a secondary (much larger) flow through the propeller that produces thrust power. The turboprop is primarily a power producer and is described in similar terms as the piston-prop. The turboprop is primarily a replacement for the piston-prop since it is capable of higher airspeeds and greater range for a given aircraft weight because of its much lighter engine weight and lower C_{D0} . Although heavier than a turbojet or a turbofan because of the propeller and gearbox, it is of the order of four times lighter than a piston-prop of the same horsepower. Furthermore, although the frontal area is somewhat larger than that of a turbojet, it is less than that of a piston-prop, and when the engine is operating, the C_{D0} is of the order of that of a turbojet, which means higher E_m 's than a piston-prop. The presence of the jet thrust, which though relatively small is essentially constant, tends to flatten the thrust curves at the higher speeds and to reduce the rate of decrease of the propulsive efficiency. The turboprop has a low specific fuel consumption, of the order of but somewhat higher than that of the piston-prop. One other major advantage over the piston-prop is its much lower maintenance costs. Although its initial cost is higher, it is a simpler engine with greater reliability, especially with the recent improvements in the gearbox.

A turbofan is a multiflow engine similar in many respects to a turboprop, except that the additional turbines directly drive a fan that resembles an axial flow compressor. Even though with very high bypass ratios the turbofan may produce more power than thrust and perform more like a turboprop than a turbojet, it is customary to describe it as though it were a turbojet.

The turbofan combines the good propulsive efficiency and high thrust at lower airspeeds of the piston-prop with the constant thrust and increasing propulsive efficiency at the higher airspeeds of the turbojet. Since the complexity and weight of the reduction gearbox and propeller governor system of the turboprop are eliminated, the turbofan is even simpler and lighter. Furthermore, the airflow through the ducted fan is not greatly affected by the airspeed, so that the decrease in propulsive efficiency at high air-

speeds is not as significant as the decrease associated with the propeller efficiency of the turboprop. Consequently, the turbofan can be used at airspeeds up to and including low supersonic airspeeds. Although the frontal area is larger than that of the turbojet, the turbofan is considerably shorter and the overall drag is not necessarily larger. The specific fuel consumption is much less than of the turbojet and, although it is more than that of the turboprop, it approaches comparable values. The turbofan is also quieter than the turbojet and much quieter than the turboprop, an advantage in these days of increasing concern with noise pollution.

Since both the turboprop and turbofan are multiflow engines, the equivalent specific fuel consumptions are combinations of hpsfc's and tsfc's and thus will vary with airspeed. Any value quoted in the literature is for a specific airspeed, which is not always given. The variation in the specific fuel consumption with airspeed is greater for the turbofan than for the turboprop.

Since turboprops and turbofans represent different combinations of the piston-prop and turbojet, their performance should fall somewhere between that of the piston-prop and that of the pure turbojet. Comparing turboprop, turbofan, and turbojet engines of comparable power (comparable gas generators), one finds that the turboprop will deliver the largest amount of thrust at the lower airspeeds, to include the aircraft standing still at the start of the takeoff run. The thrust, however, will decrease at the most rapid rate of the three as the airspeed increases and, at liftoff, will probably be less than the other two. The turbofan will produce less thrust than the turboprop at lower speeds but more than the turbojet, which not only improves the takeoff and early climb performance but also allows higher gross weights for takeoff than the turbojet. The thrust decreases with increasing airspeed but at a slower rate than does that of the turboprop because of the differences between a fan and a propeller and because of the larger jet thrust component. As the bypass ratio increases, the performance of the turbofan approaches that of the turboprop at the lower airspeeds but retains some of the characteristics of the turbojet at the higher airspeeds. The turbojet has the lowest initial thrust of the three, but the thrust essentially remains constant with airspeed.

With respect to the other aspects of performance, the turboprop is sufficiently similar to the piston-prop that it is a reasonable approximation simply to use the piston-prop equations without modification. The turbofan, however, is not necessarily as simple or as straightforward. If the bypass ratio is low, the turbojet equations can be used without modification. As the bypass ratio is increased and the ratio of power to thrust increases, the turbofan takes on more of the characteristics of the turboprop and piston-prop, particularly at the lower airspeeds. It is still possible

to use the turbojet equations with the realization that the actual low-speed values might be somewhat different. At the higher airspeeds, the turbofan will perform more like a turbojet but with a lower specific fuel consumption.

Because of its superior fuel efficiency, the turbofan quickly replaced the turbojet for subsonic aircraft, and the turboprop replaced the piston-prop in many applications (such as commuter aircraft) because of its lighter weight and higher airspeeds. In recent years, however, the smaller *regional jet* (*RJ*), with 30–100 seats, is rapidly replacing the turboprop as a feeder and short-range commercial aircraft (see Fig. 8). Although the *RJ* fuel efficiency is lower than that of the turboprop, the *RJ* has the advantages of higher cruising altitudes (above turbulence and weather), shorter flight times, less noise, and possibly higher public acceptance because of its more modern appearance.

VII. CONVENTIONAL SUBSONIC AIRCRAFT

A. The Specific Range

For an aircraft in unaccelerated level flight, the lift generated by the lifting surfaces must be equal to the weight of the aircraft, and the thrust produced by the propulsion system must be equal to the drag, which resists the movement of the aircraft through the air. For almost any aircraft, the distance it can fly with a given fuel load (its range) is an important performance and design criterion. The specific range (the distance traveled per unit weight of fuel) of a pure turbojet is given by Eq. (23) and for a piston-prop by Eq. (60). In these equations V is the true airspeed, D is the drag, E is the flight lift-to-drag ratio (the aerodynamic efficiency), c and \hat{c} are the respective specific fuel consumptions, η_p is the propeller efficiency, and k is a conversion factor from engine power to propeller thrust power.

The specific range of an aircraft is the exchange ratio between range and fuel; it has the units of miles per pound of fuel and is analogous to the mileage of an automobile, especially when expressed in miles per gallon (mpg). (The approximate weight of a gallon of jet fuel is 6.75 lb and that of a gallon of gasoline is 6 lb.) Examination of Eqs. (23) and (60) shows that whereas the mileage of a turbojet/turbofan or turboprop is directly proportional to its airspeed, that of a piston-prop or turboprop is explicitly independent of the airspeed but directly proportional to the propeller efficiency. We must not forget, however, that the aerodynamic efficiency of any aircraft is directly coupled to its airspeed. These two expressions also show that the operational economy for both types of aircraft can be improved (the mileage increased) by improving the aero-

dynamic efficiency (reducing the drag) and by reducing both the specific fuel consumption and the total weight of the aircraft. Since turbofans and turboprops combine, to varying degrees, the features of both the turboprop and the piston-prop, their performance falls somewhere between these two with the turbofan described as though it were a turbojet and the turboprop described as though it were a piston-prop.

B. Drag Reduction

Since the wing is the major source of the lift and drag of an aircraft, it can be used to discuss drag reduction options. The drag is directly proportional to the drag coefficient C_D , which can be divided into two components, the zero-lift drag coefficient C_{D0} and the drag-due-to-lift coefficient C_{Dl} . With the common assumption of a parabolic drag polar, the drag coefficient can be written as:

$$C_D = C_{D0} + KC_L^2 \quad (62)$$

where:

$$K = 1/[\pi(AR)e] = S/(\pi b^2 e) \quad (63)$$

In Eq. (4), AR is the aspect ratio of the wing, S the wing area, b the wing span, and e the Oswald span efficiency, which is of the order of unity or less. If C_{D0} incorporates the drag of other components of the aircraft (fuselage and nacelles, for example), Eq. (3) can be considered to be the drag polar of the aircraft, and the maximum lift-to-drag ratio of the aircraft (E_m) can be shown to be

$$E_m = 1/[2(KC_{D0})^{1/2}] \quad (64)$$

It should be noted that E_m is a design characteristic of a particular aircraft; its value cannot be exceeded although an aircraft can, and often does, fly at lower values. The higher the value of E_m is, the larger the mileage can be. It should also be noted that C_{D0} and K can be assumed to be constant until the airspeed reaches the drag-rise Mach number M_{DR} , where the drag for a constant C_L begins to increase because of the formation of shock waves as the transonic region is approached.

We see that the drag can be reduced, and the aerodynamic efficiency increased, by reducing C_{D0} and by increasing both the AR (by reducing S or increasing b) and the Oswald span efficiency e . The application of computational fluid dynamics (CFD), using the ever-increasing power of modern computers, to the calculation of the flow fields around the wing and fuselage has led to more efficient wings and lower-drag fuselages and nacelles. In addition to a reduction in C_{D0} , such wings (such as the supercritical wing) can either be made thicker (reducing weight) or have the sweep angle reduced (increasing b

and e) without affecting the M_{DR} . Alternatively, by maintaining the same thickness and sweep, the M_{DR} can be increased.

It has been demonstrated that boundary layer control can extend the region of the laminar flow over the wing, control the turbulence, and greatly reduce the drag. Unfortunately, the active systems used to date to remove or reduce the turbulent flow require energy and hardware and have operational problems, such as keeping the myriad suction holes of one scheme unplugged. Passive schemes using riblets or shallow (a fraction of an inch) chordwise grooves similar to those used on several racing yachts show promise and are being tested.

The Oswald span efficiency e is a measure of the aerodynamic losses associated with the vortices at the tip of a finite wing. As the aspect ratio increases, so does the span efficiency, approaching unity in the limit as the wing span becomes infinite. Furthermore, the addition of properly designed wing tips, end plates, and winglets to interfere with the vortex formation will also increase the span efficiency. In addition to increasing the Oswald span efficiency, which reduces the drag, an increased AR in itself also reduces the drag. Because of the coupling between the AR and e , they are sometimes combined into one parameter, $(AR)e$, which is called the effective aspect ratio.

A large AR , which is obviously desirable from the viewpoint of reduced drag, calls for a large wing span and/or a small wing area. There are structural and weight limitations associated with a large wing span as well as such practical considerations as the size of hangar doors and the space required for using loading docks at passenger terminals. As the wing area is decreased, the wing loading increases, calling for larger takeoff and landing airspeeds, thus requiring longer runways. In most cases, the maximum allowable wing loading (smallest wing area) is established by the maximum allowable approach airspeed, which in turn is determined by the distance of the landing run.

Equation (62) shows that, in addition to changing physical characteristics, the wing drag can also be reduced by decreasing the C_L of the wing. With a horizontal tail surface and the center of gravity (cg) of the aircraft behind the aerodynamic center (ac) of the wing, as shown in Fig. 7(a), the sum of the wing lift and the tail lift must equal the weight of the aircraft in level (cruise) flight, and the sum of the pitching moments about the cg must be equal to zero. Consequently, the tail lift will be positive and the lift to be generated by the wing need only be equal to the weight minus the tail lift, requiring a smaller C_L and thus reducing the wing drag. If the cg is moved further aft, the required wing lift decreases, further decreasing the C_L and the drag of the wing. Since the required tail lift will increase appropriately so as to keep the pitching moments

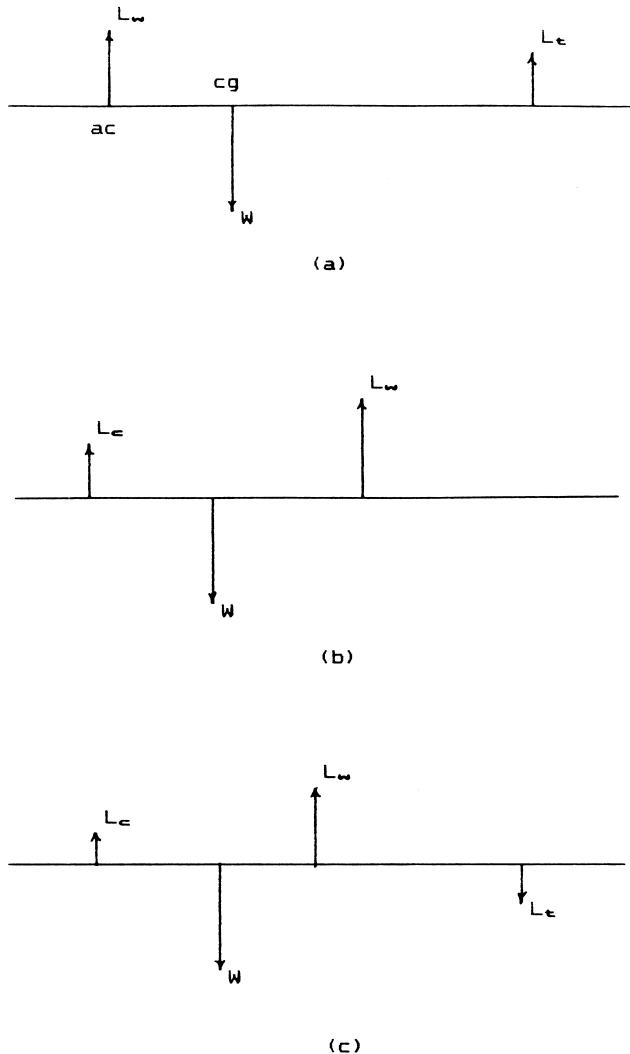


FIGURE 7 Possible wing–horizontal stabilizer combinations: (a) two-surface—wing and tail; (b) two-surface—wing and canard; and (c) three-surface—wing, tail, and canard.

equal to zero, the C_L and drag of the tail will increase. However, the increase in tail drag will be much smaller than the decrease in the wing drag because of the smaller tail area. The latest versions of the Airbus and Boeing 747 are being built with a fuel tank in the tail so that the cg can be shifted farther to the rear during cruise than would be possible with the normal wing tanks. An added benefit of having the cg behind the ac of the wing is the fact that the wing should stall first (since the angle of attack of the tail is smaller than that of the wing); when it does, the positive tail lift will generate a nose-down pitching moment to initiate a stall recovery.

Canards, horizontal lifting surfaces placed ahead of the wing, are beginning to reappear (the Wright Flyer used a canard), but there is some question as to whether they are



FIGURE 8 The Canadair Regional Jet, CRJ 200. A 50-passenger, Mach 0.81 transport with two turbofans of 8700-lb thrust, a gross weight of 53,000 lb, a wing area of 520 ft^2 ($W/S = 102 \text{ lb}/\text{ft}^2$), and a range of 2300 miles. [Courtesy of Bombardier Aerospace.]

primarily a matter of styling or a special-purpose solution rather than a general improvement over the horizontal tail. In any event, with a canard, the cg should be ahead of the ac of the wing, as shown in Fig. 7(b), so that the beneficial effects on the wing lift and drag may be realized; it may be more difficult to shift the cg with a canard. In this configuration, it is important that the canard stall first. There is also at least one three-surface configuration being built; Fig. 7(c) shows a three-surface combination with the cg ahead of the ac.

The drag of the tail can be further reduced by elimination of the tail, resulting in a “flying wing,” as with the B-2B Stealth Bomber (see Fig. 11) and the *blended wing body* (BWB) design studies for transports now underway. (BWB is another way of saying flying wing.). Removal of the *empennage* (the tail) also reduces the overall weight of the aircraft and is possible with the advances in automatic flight control systems.

C. Propulsive Efficiency

Although the piston-prop has the lowest specific fuel consumption of all the air breathers, it is also the heaviest and has the largest drag, and its best-range airspeed is 24% lower than that of a comparable turbojet. Furthermore,

the propeller efficiency drops off sharply at relatively low Mach numbers as the propeller tip speeds approach the sonic velocity. The turbojet produces its thrust by expanding all of the turbine exhaust gases through a nozzle, and it has the highest specific fuel consumption but is the lightest. It also has the fewest moving parts and the lowest drag, and its propulsive efficiency improves with airspeed. The turbofan and the turboprop are basically turbojets in which part of the exhaust gases is used to drive an ungeared multibladed ducted fan (the turbofan) or a propeller connected to the turbine drive shaft through a gearbox (the turboprop). The ratio of the mass of the cold air passing through the fan (or the propeller) to the mass of the hot air passing through the burners and turbines is called the bypass ratio. If the bypass ratio is zero, the turbofan becomes a pure turbojet; current operational bypass ratios are of the order of 5–6. As the bypass ratio increases, specific fuel consumption decreases and the turbofan begins to take on the characteristics of a turboprop with the exception that the ducted fan efficiency is essentially independent of the airspeed. Although it is not customary to refer to the bypass ratio of turboprops, it is of the order of 50.

Although turboprops have a specific fuel consumption that approaches that of the piston-prop and are four times lighter, their cruise speed has been limited by the

degradation in the propeller efficiency at higher airspeeds. The fuel shortage in the mid-1970s and the subsequent increase in prices sparked the development of two ultra-high bypass (UHB) engines, one known as a propfan and the other as an unducted fan (UDF), which in appearance resemble each other. They each have 2 rows of counter-rotating blades (6 or more in each row) that are approximately 12 ft in diameter; the blades do not resemble conventional propeller blades, being highly swept and having variable camber. The propfan is essentially a turboprop with a gearbox that reduces the high rpm of the turbine to that of conventional propellers; the rpm and the pitch of the blades can be varied. The unducted fan, on the other hand, is essentially a turbofan with a bypass ratio of the order of 28; it has no gearbox and operates at a constant rpm (that of the turbine), although the pitch of the blades can be varied through feathering and into reverse. With the duct removed, it is necessary to give greater consideration to the fan efficiency.

Although the propeller efficiency has been extended, it still drops off at the higher subsonic Mach numbers. Furthermore, the best-range airspeed of aircraft using both of these UHB engines is still lower than that of the pure turbojet and turbofan so that at competitive airspeeds the UHB lift-to-drag ratio will also be lower. In spite of these off-design point degradations at the high subsonic Mach numbers of current airliners, the specific range of UHB-equipped aircraft is sufficiently higher to make such an aircraft commercially attractive at Mach 0.8.

Although both types of engines were flown on demonstrator aircraft and design studies for commercial aircraft with UHB engines were started, current efforts seem to have ceased.

The thermal efficiency of gas turbine engines is strongly dependent on the turbine inlet temperature, which is primarily limited by the physical properties of the turbine blades. There are ongoing investigations into new materials and alloys, into techniques for forming turbine blades and fastening them to the hub, and for controlling the clearance between the blade tips and the hubs. As the turbine inlet temperature increases, so does the exhaust temperature, with a subsequent increase in acoustic noise, which must be considered in light of the drive for quieter engines.

In concluding this section, mention should be made of the increasing use of derated turboprops to replace piston (reciprocating) engines; derating means that the propeller selected cannot absorb all of the sea-level power developed by the engine. As the engine shaft power decreases with altitude, the propeller thrust power will remain constant until the maximum engine power available decreases to that value. The altitude at which the engine power and propeller thrust power are equal is analogous to the critical altitude of a turbocharged piston-prop and can be as high as 20,000 ft; the derated turboprop behaves like a

turbocharged piston-prop but with a much lighter weight and less complexity.

D. Weight Reduction

A 5% reduction in weight means a 5% increase in the specific range and generally a reduction in the initial cost of the aircraft. Consequently, weight reduction, along with increasing propulsion performance, is always an active program.

Since the wing is the heaviest component of an aircraft, the decreases in the wing area being made to reduce the drag also reduce the weight of the aircraft. In addition, the use of active flight controls to shift air loads from the outboard region of the wing toward the wing roots means a lighter wing for a given wing span or a higher aspect ratio wing for the same weight. Active controls also allow stability requirements to be relaxed, permitting smaller horizontal tail (and canard) surfaces and smaller tail lengths, thus reducing both the weight and the tail drag.

New aluminum alloys can cut weight with a slight increase in cost, but the use of advanced composite materials, such as graphite-fiber laminates, shows the greatest promise for major reductions in structural weight. Over the past years, there has been an increasing use of composites for secondary structures (rudders, speed brakes, flaps), but now primary structures (wings, fuselages, empennages) are being made from composites. Approximately 80% of the structure of the tiltrotor aircraft (including the rotor blades) to be discussed in Section VIII is fabricated from composite materials with an estimated weight savings of 25%. Although the current overall cost of using composites is approximately the same as that of conventional materials, the labor costs are less. With reductions in the cost of the composite materials and improvements in the strength-to-weight ratios of the fibers and laminates, the savings in both cost and weight should increase. Additional advantages of composite structures are greater corrosion resistance and ease of repair.

The weight of the gas turbine engines continues to be reduced, thus increasing the payload ratio of an aircraft. Other areas of future weight reduction are the use of high-pressure (8000 psi) hydraulic systems and the replacement of hydraulic actuators by electromechanical actuators and hydraulic lines by electrical wires and optical fibers.

VIII. VERTICAL AND SHORT TAKEOFF AND LANDING AIRCRAFT

The horizontal takeoff and landing runs of a conventional fixed-wing aircraft are directly proportional to its wing loading. Consequently, as the wing loading increases to

improve the performance (block time and range), so does the required runway length; as the traffic at a particular airport increases, so does the need for additional runways and additional land. As a result, airports for metropolitan areas are generally very large and located at some distance from the heart of the city, creating traffic and parking problems and increasing the overall travel time.

The takeoff run is also inversely proportional to the thrust-to-weight (T/W) ratio and to the lift coefficient C_L used for takeoff. Increasing the thrust, however, requires heavier engines, which increase the weight of the aircraft and thus reduce the specific range of the aircraft. In turn, more fuel is needed, which also increases the weight and requires larger engines. In addition, as the weight increases, the wing area needs to be increased if the original wing loading is to be maintained; then, the weight will increase and the cycle begins again. With respect to reducing the takeoff run by increasing the C_L beyond the values currently achieved with leading-edge and Fowler-type trailing-edge flaps, the high-lift flap devices and techniques currently under study would add to the weight of the aircraft and compromise the aerodynamic efficiency. Furthermore, for high wing-loading aircraft, there is a trade-off between the decrease in the ground run (favorable) and the increase in drag (unfavorable); aircraft usually use less flaps for takeoff than for landing, when both high drag (for steeper descents) and low airspeed (for shorter landing runs) are desirable.

Although the prospects of reducing the runway requirements for long-range aircraft are not encouraging, the same is not true for short- and medium-range transports used for travel between smaller towns and cities, where long runways are not available, and the large hub airports for interface with long-range aircraft and for travel between city-pairs that are not great distances apart, such as the New York–Washington or San Francisco–Los Angeles city-pairs. Surveys have shown that the average range for such city-pairs within the United States is of the order of 200 miles.

With such ranges, the airspeed (and block time) is not as important as the runway length reduction, and STOL (short takeoff and landing) aircraft begin to look attractive. They are characterized by lower wing loadings and a large C_L , resulting in low takeoff and landing airspeeds, and possibly an increased T/W (with a lower range requirement, the additional wing, engine, and flap weights can be accommodated by reducing the fuel weight).

With half the wing loading and double the takeoff C_L , for example, the takeoff run will be approximately one-fourth of the original distance; doubling the T/W ratio will further reduce the takeoff run to one-eighth. Furthermore, the climb-outs and descents of STOL aircraft will be much steeper, so that departure and approach patterns will

be much smaller and obstacle clearance (buildings on the flight paths, for example) will be less of a problem. With these reduced requirements, there are plans for STOL runways at the hub airports, so located that the traffic patterns do not interfere with those of the main runways, and for exclusively STOL airports to be located closer to the population centers. Unfortunately, these plans do call for additional land, even though the amount is less, in areas where land is unavailable or scarce and very expensive, and little has been done to produce a true STOL aircraft. However, work on high C_L technology continues with emphasis on maintaining attached flow over large trailing-edge flaps by blowing, either by using power from the engines or by mounting the engines above the wing and using the exhaust or slip-stream directly.

V/STOL aircraft (aircraft that can take off and land vertically and have an inherent STOL capability) appear to be the appropriate solution for the feeder and intercity missions; they are also attractive to the military for tactical use from unprepared sites and from ships. Such aircraft, however, require a vertical T/W ratio greater than unity. Helicopters can, and do, meet this requirement and are numerous. Unfortunately, they also have high drag and low aerodynamic efficiencies (lift-to-drag ratios of the order of three), are noisy, and are limited to short-range missions and to airspeeds less than 200 mph and altitudes less than 20,000 ft. When the weight of a helicopter exceeds its vertical lift capability, takeoff can be accomplished by using a short run, giving the helicopter a STOL capability.

Although useful, the helicopter is not the solution to the feeder and intercity problem. What is needed is an aircraft that can take off and land vertically and cruise as a conventional aircraft. One early and unconventional approach was to stand a pistonprop aircraft (the POGO) on its tail for takeoff and landing and then rotate it to the horizontal for cruise; this was not a satisfactory solution. Other approaches investigated included using lift fans in the wings for vertical flight, tilting the wing so that the engines were vertical for takeoff and landing, and swivelling the nozzles of a jet engine so that the thrust would be vertical for takeoff and landing. This last approach was adopted and is being used in a military combat aircraft, the Harrier, which is a tactical strike aircraft operating from both land and ships. As with the helicopter, larger payloads that exceed the vertical lift capability can be handled by operating the aircraft in a STOL mode, sometimes with an inclined ramp resembling a ski jump. Unfortunately, the weight of the large jet engine(s) required to give a T/W ratio greater than unity results in an aircraft with a low payload capability and a short range and, consequently, no real commercial application.

Another approach investigated was the tiltrotor concept in which large propellers (rotors) are located at each tip of a



FIGURE 9 The Bell Boeing V-22 Osprey landing on the deck of an assault carrier. A military tiltrotor V_{STOL} (vertical and short takeoff and landing) transport with a capacity of 24 combat troops, two 6150-SHP turboshaft engines, a maximum vertical takeoff weight of 47,000 lb, and a maximum airspeed of 315 mph. [Courtesy of Bell Boeing.]

fixed wing and are tilted from the vertical to the horizontal for cruise to combine the capability of the helicopter with that of the fixed-wing aircraft. After proving the soundness of this approach with a demonstrator aircraft, the XV-15, the U.S. Department of Defense authorized the development and production of the V-22 Osprey Tiltrotor aircraft (see Fig. 9). This is a 50,000-lb aircraft with two turboshaft engines, each in its own tiltable pod driving a three-bladed 38-ft-diameter, rotor. It is a unique aircraft in many ways, including the use of composites for the rotors and the majority of the structure, and should truly merge the performance of a helicopter with that of a turboprop. It is expected to have a maximum airspeed in excess of 300 mph, a ceiling of the order of 28,000 ft, and a maximum specific range approaching that of the turboprop but at a higher cruise airspeed. Although this is a military aircraft with a 6000-lb payload and a range of the order of 1000 miles, the commercial potential has been recognized and a 40-passenger conversion should pose no problems. It is of interest to note that the XV-15 demonstrator aircraft flew from Battery Park in lower Manhattan to the Pentagon in Washington, D.C., in 40 min. The tiltrotor may well be the solution to the intercity and feeder problem.

IX. MILITARY AIRCRAFT

In addition to the tiltrotor of the previous section, the military is interested in increasing the agility of its fighter aircraft; agility can be thought of as comprising maneuver-

ability (turning, diving, climbing), controllability, and acceleration. After detection and identification, aerial combat between individual aircraft may well start with the launch of air-to-air missiles at high supersonic Mach numbers (Mach 2.0 to Mach 3.0) and at high altitudes (50,000–60,000 ft), with the aircraft functioning simply as launch platforms for the missiles. Some interceptor aircraft in the past were designed primarily for that function, with limited maneuverability and even without guns. Without a missile kill, the engagement becomes a dogfight, and survivability and victory depend upon the maneuverability of the aircraft and the skill of the pilot. Since the maneuverability of an aircraft improves as airspeed and altitude decrease, as the engagement continues, the aircraft will slow down and descend until the low-speed, low-altitude combat arena is reached (or one of the aircraft is destroyed). Although the maneuverability is highest at sea level, the lower limit to the combat altitude is the minimum altitude required for pull-up from an evasive maneuver.

The maneuvering ability of an aircraft can be described in terms of its maximum turning rate (fastest turn) and its minimum turning radius (tightest turn). For best turning performance, both maneuvers theoretically call for a large T/W ratio, a low wing loading, a large aspect ratio, a low C_{D0} , a low altitude, and low airspeeds. Only the high AR is missing in modern fighters because of the structural, weight, and aerodynamic considerations involved in high-*g* and supersonic flight. Furthermore, increasing the wing area so as to decrease the wing loading requires a larger wing span [$b = (S \times AR)^{1/2}$] to maintain the same

AR , also increasing the weight and structural problems. Although steady-state fastest and tightest turns are not necessarily high- g maneuvers, there are many times when the aircraft will be put in the position of turning with high- g forces. Consequently, modern fighter aircraft rely on lower wing loadings, lower C_{D0} 's, and higher T/W's for their maneuverability and accept the penalties associated with lower AR 's and lower lift-to-drag ratios. High T/W's also provide sufficient acceleration for an aircraft to escape and regain airspeed and altitude before rejoining battle.

Although the design characteristics for fastest and tightest turn are similar if not identical, the two types of turns do differ. First of all, the load factors (g 's) are different; the fastest-turn load factor is proportional to the square root of the maximum attainable load factor and has no theoretical upper limit, whereas the tightest-turn load factor cannot exceed 1.4 g 's. Then, although theoretical airspeeds are low (often lower than the stall speed of the aircraft), the tightest-turn airspeeds are even lower than those for the fastest turn.

Since, in both cases, the turning performance improves as the airspeed is decreased, it seems obvious that every effort should be made to reduce the stall speed by increasing the maximum lift coefficient and the high angle of attack aerodynamic performance. This has been the major objective of various tests of canards, strakes, and variable camber (mission adaptive) wings and of such devices as drooped leading edges and thrust deflectors, which are designed to improve stability and controllability at these high angles of attacks and low airspeeds. In addition, there have been tests of vertical and canted canards, which, when used in conjunction with other control surfaces in a fly-by-wire system, can translate the aircraft horizontally and vertically without rotation. They can rotate the aircraft about the vertical axis without banking or turning.

The X-31 Enhanced Maneuverability Aircraft Program was a joint U.S./German program with two prototype aircraft built by Rockwell/MBB. These aircraft use a canard and a variable camber wing along with a thrust-vectoring capability. The vectored thrust will not only improve the turning performance but will also so improve the controllability at high angles of attack that the aircraft can break the stall barrier to fly in the poststall region at angles of attack larger than the normal stall angle and at airspeeds lower than the stall speed.

Since its targets are usually located deep inside the enemy country, a strategic bomber must survive or evade attacks by enemy surface-to-air defenses and interceptor aircraft if it is to accomplish its mission and then return. During World War II, this penetration, as it is called, was carried out at high (approaching the ceiling of the aircraft) altitudes above or at the extreme end of the range of anti-aircraft guns with defense against enemy aircraft by escort

aircraft (within their limited-by-fuel range) and by guns aboard the bomber. After the war, the B-52, a subsonic jet bomber, was designed to penetrate in such a manner. As antiaircraft guns were replaced by long-range surface-to-air missiles with various types of guidance, many with terminal guidance systems, and interceptor aircraft were equipped with air-to-air missiles, it became apparent that the B-52 was very exposed and vulnerable at high altitudes. Consequently, its mission was changed to a low-level (on-the-deck) penetration where it would be difficult, if not impossible, to track the B-52 with ground-based radar (although it would be exposed to downward-looking radar), where it could not be reached by the long-range missiles and where the time of exposure (the time to cross from horizon to horizon) to specific short-range missiles and other weapons would be short.

The X-31 Enhanced Maneuverability Fighter Program was a joint U.S./German program with two prototypes built by Rockwell/MBB. These aircraft used moveable canards and strakes, a variable camber wing, and thrust vectoring to improve both the turning performance and the controllability at high angles of attack and low airspeeds (the combat arena for fighter aircraft). It was possible with the X-31 to break the stall barrier and to fly in the poststall region at angles of attack larger than the normal stall angle and at airspeeds lower than the conventional stall speed. The results of this program have been introduced into the F-22 Raptor (see Fig. 3), an all-purpose, highly maneuverable fighter with supercruise capability, vectored nozzles, a large complement of weapons (including a 20-mm cannon) in internal bays, and stealth characteristics.

The B-1B Lancer strategic bomber (see Fig. 10) also penetrates at low level at a high subsonic speed, using its supersonic dash capability to reduce the time required to reach the penetration area. The B-1B has a low wing with variable sweep; the leading edge can be varied from 15 to 67.5° , reducing the wingspan from 137 ft to 78 ft. The aircraft is almost 150 ft in length.

The B-1B has four very low bypass ratio turbofan engines, each with approximately 31,000 lb of unaugmented thrust: afterburners are used for supersonic flight with a maximum Mach number of the order of 1.25. It has a four-man crew and a low radar cross section and is equipped with sophisticated offensive avionics and terrain following systems. The structure, of special aluminum alloys and titanium, is hardened to sustain nuclear blasts and overpressures. The maximum gross weight is approximately 500,000 lb and the unrefueled range is of the order of 7500 miles. There are three internal (rotary) weapons bays to carry missiles and bombs.

Although vulnerability is reduced, low-level penetration has its own problems. Terrain clearance and obstacle avoidance, either manually or automatically, are



FIGURE 10 The B-1B Lancer, originally built by Rockwell International. A low-level penetration strategic bomber with a Mach 1.2 capability, a crew of four, a swing wing, four low-bypass turbofans of the 31,000-lb (unaugmented) thrust class, and a maximum gross weight of 470,000 lb. [Courtesy of Boeing.]

not simple tasks. Furthermore, at these low altitudes and high airspeeds, the dynamic pressure is very high, which, when combined with the maneuvers associated with terrain clearance and avoidance and low-altitude turbulence, imposes large structural loads on the aircraft. For example, the B-52s have had their main wing spars replaced and have been reskinned at least once. Cruise missiles carried by the penetrating bomber can be launched at low altitudes but standoff missiles and thermonuclear (TN) bombs will require a pop-up maneuver of some kind at the end of the penetration.

In spite of the reasons given above for low-altitude penetration, the newest bomber, the B-2 Spirit, is designed to penetrate at higher altitudes at a high subsonic airspeed; it does not have a supersonic capability. The B-2 is popularly referred to as the STEALTH bomber because it is designed to be as undetectable as possible by dramatically reducing both its radar cross section (RCS) and its infrared (IR) signature. If the bomber is detected, it is expected that the reduced radar and IR targets, along with chaff and electronic and heat decoys, will prevent effective tracking and lock-on of surface and air-launched missiles. The combination of a low-level penetration by the B-1B and a high-altitude stealthy penetration by the B-2 are expected to saturate and confuse existing enemy radars and defenses and require the deployment of additional and new defenses.

The B-2, being built by Northrop/Boeing, is a flying wing (see Fig. 11). It is a descendant of the XB-49, which

was designed by Jack Northrop and flown in the late 1940s. (The XB-49 had stability and control problems and a limited cg shift, problems that can be handled by modern flight control and stability augmentation systems.) The B-2 has a crew of two and is powered by four engines. Its approximate dimensions are a height of 17 ft, a length of 69 ft, and a wing span of 172 ft, so that it is only slightly longer than an F-15 fighter but has a wing span approaching that of a B-52.

Among the techniques used to reduce the RCS are the absence of vertical stabilizers to eliminate radar reflections, the use of composites for the airframe to absorb microwaves, the use of radar-absorbing blankets around the engines and weapons bays, and the use of S-shaped engine air inlets to prevent radar from reaching and detecting the compressor blades. It is believed that the purpose of the saw-toothed trailing edge of the wing is to permit the engine exhausts to be placed against the trailing edge so as to reduce the IR signature. Since hot exhaust gases can be easily detected from long distances, they will probably be cooled and diluted by the addition of ambient air. There obviously will be no external stores whatsoever with all of its 45,000 lb of payload being carried internally, probably with a rotary carrier of the type used in the B-1B. Current planning calls for a fleet of 132 aircraft with an estimated cost on the order of \$60 billion.

There is also a STEALTH fighter, the F-117A Nighthawk, built by the Lockheed Martin Corporation (see Fig. 12). Although the aircraft became operational



FIGURE 11 The B-2 Spirit (the STEALTH Bomber), originally manufactured by the Northrop Corporation. A high-subsonic-flying wing with multirole missions, a crew of two, four high-bypass-ratio turbofans of the 18,000-lb class, and a typical maximum weight of 335,000 lb. [Courtesy of the U.S. Air Force.]

in 1983, it was not officially acknowledged until 1988. It is a single-place, subsonic, twin-engine aircraft with a delta wing, without a horizontal stabilizer but with double canted, vertical stabilizers. It is a penetration strike fighter to be used against radar and missile installations and other special targets, such as command posts. It was used in the Gulf War.

X. ADVANCED SUPERSONIC COMMERCIAL AIRCRAFT

The actual and projected increase in traffic among the Pacific Rim countries has sparked a renewed interest in an advanced supersonic transport (SST). Currently, a nonstop flight from Los Angeles to Tokyo takes 11 hr or more in a high subsonic Mach number transport; a Mach 2.7 aircraft would reduce that time to less than 4 hr. Such an aircraft would have to be economically competitive, amortizing its development and production costs as well as meeting its operating costs and providing a reasonable profit by charging no more than first-class fares. The Concorde is apparently showing a profit but it is not attempting to recoup its development and production costs and is charging premium fares for a limited group of passengers.

Mach 3 seems to be a reasonable upper limit for an advanced SST, particularly when thinking about a range of the order of 7500 miles (6500 nautical miles). At higher Mach numbers, the higher skin temperatures lead to requirements for large amounts of insulation (and possibly no fuel in the wings) and active cooling, as well as the possibility of being forced to switch to more exotic materials. The higher skin temperatures may also require a cooling-off period after landing that could delay both the unloading and turn-around time. Furthermore, turbojets reach their operating limits at about Mach 3, so new combined-cycle engines (turbo-ramjets) would be needed. Finally, the time spent in acceleration, deceleration, and ground operations, rather than actual cruise time, would dominate the block time.

An important indicator in determining and comparing operational efficiencies of passenger transports is the value of the seat-miles per gallon of fuel (seat-miles/gal) parameter, which is simply the mileage of Eq. (23) multiplied by the number of passengers.

A current Mach 0.85 transport with 450 passengers, a lift-to-drag ratio (E) of 17.7, and a c of 0.6 and weighing 800,000 lb has a value of 63 seat-miles/gal. On the other hand, an SST with a cruise Mach number of 2.7, 250 passengers, an E of 10, and a c of 1.5 and also



FIGURE 12 The F-117A Nighthawk, a high-subsonic, defense suppression aircraft with a crew of one, low detectability, internal weapons storage, two turbofan engines of the 12,000-lb thrust class, and a weight of 52,000 lb. [Courtesy of Lockheed Martin.]

weighing 800,000 lb has a seat-miles/gal value of the order of 25, which is approximately 40% of that of the subsonic aircraft. At first glance, the SST hardly seems competitive; however, since its block time is significantly lower, its utilization factor should be much higher. If the SST can make 2.5 flights for each flight of the subsonic aircraft, then the seat-miles/gal figures become comparable and the competitive position of the SST becomes stronger.

Without going into details, the supersonic lift-to-drag ratio of the SST is considerably lower than that of the subsonic aircraft mainly because of the presence of wave drag, and the specific fuel consumption is much higher because high bypass ratio turbofans cannot be used. The engines being considered are essentially turbojets (with bypass ratios on the order of 0.25) with afterburning for passage through the transonic region.

The appearance and characteristics of an advanced SST will also be different. One possible configuration shows a slender, blended wing-body configuration, a double-swept wing, and curved wing tips, all designed to reduce the drag. The aspect ratio will be lower (on the order of two) to increase the slenderness ratio, but, since the wing

area will be larger (approximately double), the reduction in wing span will not be as large as expected; the wing loading will be about half that of a comparable subsonic aircraft. Since the body will be very slender, the flight crew may well be located a considerable distance behind the nose without any windows; in fact, there may not be any windows whatsoever so as to simplify the structural design and environmental control of the cabin in the presence of exterior skin temperatures on the order of 550°F during cruise. External visibility would be provided by individual video systems.

Although there is interest in an advanced SST in the United States and abroad, it is difficult to make a strong case for such aircraft; current activity is limited to design studies with no plans for actually developing supersonic transports. Other than the Concorde and some research vehicles, supersonic flight is limited to military aircraft: the B-1B bomber, which makes a supersonic dash to the start of its low-level subsonic penetration and various fighter aircraft that use such speeds to reach a combat area, make an interception, and launch air-to-air missiles. It should be realized that if these missiles fail to destroy their targets, one-on-one aerial combat using guns

and cannons quickly becomes subsonic at lower altitudes, approaching the ground in the limit.

XI. HYPERSONIC AIRCRAFT AND THE NATIONAL AEROSPACE PLANE

Entering and leaving space using rocket engines only is a complicated and expensive process, requiring the vertical launch of heavy vehicles with two or more stages and with structures that cannot withstand side forces. Rocket engines, although light in weight, are not very efficient and cannot use oxygen from the atmosphere, so that launch vehicles using rocket engines must carry large quantities of both fuel and oxidizer. For example, 75% of the space shuttle's lift-off weight of 4.4 million pounds is propellant and 83% of that is oxygen. As one author wrote, "A launch vehicle carrying oxygen through the atmosphere is like a fish in the ocean carrying a canteen of water."

With the advent of space travel, there has been a fascination with the idea of an *aerospace plane* for both civilian and military use, a vehicle that would take off and land horizontally (CTOL) and have a single-stage-to-orbit (SSTO) capability. The same knowledge and technologies could be used for a family of *hypersonic transports* that could cruise at hypersonic velocities (Mach 5 or higher) at altitudes between 80,000 and 150,000 ft. Such aircraft could be used as civilian and military transports (the Orient Express) or as a recoverable first stage for space launches. The Aerospace Plane would use combinations of air-breathing engines to accelerate the aircraft within the sensible atmosphere (up to about 45 miles) to as high a Mach number as possible and then switch to rocket engines for the final acceleration and for flight outside of the atmosphere. (The National Aerospace Plane [NASP] program, which with its X-30 test vehicle was directed toward the development of an Mach 25 aircraft, has been canceled.)

A major element of an aerospace plane or hypersonic transport is the propulsion system, particularly the air breathers. Air breathers not only use oxygen from the atmosphere but are more efficient than rocket engines and are much more complex. However, the specific impulse of an air breather can be an order of magnitude (ten times) greater than that of rocket engines at the lower Mach numbers and two to three times higher at hypersonic Mach numbers. This greater propulsive efficiency can improve the *payload ratio*, which is the mass placed in orbit for a given gross liftoff weight (GLOW), by factors of three to five.

Unfortunately, no one air-breathing engine can span this wide spectrum of airspeeds. High specific impulse turbojets can accelerate the aircraft to Mach 2 or 3, at which point a ramjet would take over. Rather than using different

engines, the two might be combined into a single engine, the air turbo-ramjet (ATR) in which the engine compresses the intake air mechanically at the lower Mach numbers and with ram pressure at the higher Mach numbers with the incoming air bypassing the compressors and entering the combustion chambers directly. In both the turbojet and ramjet, combustion takes place with the intake air slowed down to subsonic velocities. At about Mach 6, subsonic combustion is no longer feasible since the increases in temperature associated with slowing the air down are too high, even for nonmoving engine parts. This is the time to switch to a supersonic combustion ramjet (the scramjet) in which the intake air remains supersonic; engine heating is reduced since less of the kinetic energy of the air is transformed to thermal energy. There is a problem associated with supersonic combustion, the problem of sustaining the combustion process, which is analogous to keeping a match lit in a hurricane. A fuel particle at Mach 7 has less than a millisecond to ignite before being swept out the exhaust, and the traditional hydrocarbons (jet fuel) are unfortunately too slow to ignite. However, hydrogen, which has a substantially reduced ignition time, should be satisfactory, but a switch in propellants, unless hydrogen is used in the ATR, would also be required. The short mixing time available in the scramjet requires that the hydrogen be injected from struts that span the air inlet. These injector struts will be subjected to formidable structural loads from both the shock waves in the engine flow and the differential thermal stresses from the -375°F hydrogen within the strut and the 1700°F air without.

An important, and as yet unanswered, question is the upper limit for efficient scramjet operation; current estimates range from Mach 10 to Mach 25 (orbital velocity). In any event, rocket engines would have to be used outside the atmosphere. The exact speed at which scramjets become ineffective or inefficient may determine the feasibility of an aerospace plane; the earlier the switch from scramjet to rocket, the larger the amount of liquid oxygen to be carried and the heavier the vehicle. Some skeptics say that a switch around Mach 17 or higher is needed if the takeoff size and weight are to be of the same order of magnitude as that of a conventional aircraft.

Since hydrogen will be required for the scramjets and is the logical fuel for the rocket engines, it seems wise to have a single fuel and use hydrogen for the turbojets and ramjets (or the air turbo-ramjets). Not only is hydrogen a satisfactory fuel for these engines (the Russians are currently flight testing the TU 155 with hydrogen-fueled engines), it also has other advantages. It is cleaner, with only steam as the combustion product, and, pound for pound, it has three times the energy of jet fuel. Its disadvantages are that its lower density (a fifth of that of jet fuel) requires larger tanks and, since it is cryogenic, it requires special

handling, equipment, and tanks. At present, liquid hydrogen is usually obtained from the liquefaction of natural gas and is expensive (about three times the cost of jet fuel).

The propulsion system must be carefully integrated with the airframe inasmuch as the airframe forms part of the propulsion system. The forebody of the vehicle provides the surface to shape and compress the shock wave for proper inlet entrance, and the afterbody provides the expansion surface for the nozzle. In other words, there will be no bolt-on engines. Highly sophisticated CFD will be required to accomplish the propulsion-airframe integration as well as the blended wing-body configuration. In addition, CFD will be vital to the prediction of performance and behavior at speeds and altitudes that cannot be reproduced in ground based wind tunnels.

Other areas of technology essential to the success of the aerospace plane and hypersonic aircraft are those of advanced materials, structures, and thermal control. Long-life composite structures may be needed for the liquid-hydrogen fuel tank, along with fully reusable, high-temperature, lightweight materials for the engines and airframe. Active cooling will be required for high heat flux areas, such as the nose, leading edges, and probably for the engines.

There are those who doubt the economic soundness of such a family of vehicles, who feel that the projected traffic and utilization do not warrant its development and will not provide a suitable payback, and who feel that the estimated development cost of \$3 billion is grossly underestimated. This program may well be justified only on the basis of technology development and demonstration that may be beneficial for other applications.

An aircraft is only one element of the complex air transportation system which, in addition to the aircraft, comprises: the airways and air traffic control; the airports, with their runways, taxiways, ground control, terminal buildings, and loading facilities; and other supporting facilities, such as parking lots, access roads both on and to the airports, and shuttle services. As the number of aircraft, flights, and passengers continues to increase, so does the congestion in all aspects of the transportation system, leading to increasing delays, frustration, and possibly gridlock. In an effort to reduce airways congestion, serious consideration is being given to the concept of *free flight*, whereby each aircraft is given its choice of how to proceed from one airport to another without traveling a specified and defined airway. Obviously, such free flight requires adequate navigation and control systems, both autonomous and external to the aircraft, along with proximity warning and collision avoidance systems. Although free flight might free up the airways, the airport nodes still exist and are needed for landings and takeoffs.

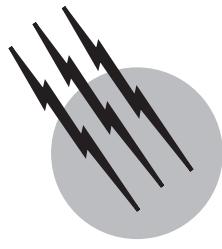
There has been no mention in this article of a growing class of airborne vehicles that do not have a pilot on board. Such a vehicle is referred to as an *unmanned aerial vehicle* (UAV) and one definition is “a powered aerial vehicle that does not carry a human operator, uses aerodynamic forces to provide aerodynamic lift, can fly autonomously or be piloted remotely, can be expendable or recoverable, and can carry a lethal or nonlethal payload.” Ballistic or semibalistic missiles and artillery projectiles are not considered to be UAVs. The development of UAVs was spurred by the need of the military for airborne intelligence, reconnaissance, and surveillance (the ISR mission) without the costs, risks, and other limitations associated with the use of human crews. Without the need to provide space and support for a human crew, the size and weight of comparable unmanned vehicles can be significantly reduced and mission duration can be increased. UAVs come in all sizes and configurations, ranging from hand-launched versions with limited payload, range, and duration to a high-altitude, long-endurance UAV flying at altitudes up to 65,000 ft with an endurance of 40 hr, a payload of 1800 lb, a weight of 26,000 lb, and a wing span of 113 ft. In addition to military missions, UAVs have a large potential for civilian uses—for example, serving as communications relays in lieu of geocentric satellites. Whatever the future of UAVs might be, they must be compatible with and safely share the air space with existing traffic.

SEE ALSO THE FOLLOWING ARTICLES

AIRCRAFT SPEED AND ALTITUDE • AIRPLANES, LIGHT • COMPUTATIONAL AERODYNAMICS • FLIGHT (AERODYNAMICS) • JET AND GAS TURBINE ENGINES • V/STOL AIRPLANES

BIBLIOGRAPHY

- “All the World’s Aircraft,” Jane’s Publishing Co., London.
- “All the World’s Uninhabited Aerial Vehicles and Targets,” Jane’s Publishing Co., London.
- Anderson, J. D., Jr. (1999). “Aircraft Performance and Design,” WCB/McGraw-Hill, Boston, MA.
- Hale, F. J. (1984). “Introduction to Aircraft Performance, Selection, and Design,” Wiley, New York.
- Tennekes, H. (1997). “The Simple Science of Flight From Insects to Jumbo Jets,” MIT Press, Cambridge, MA.
- “Aerospace America,” American Institute of Aeronautics and Astronautics, Reston, VA.
- “Aerospace Engineering,” SAE International, Warrendale, PA.
- “Aerospace International,” Royal Aeronautical Society, London.
- “Air and Space,” Smithsonian Institution, Washington, D.C.
- “Aviation Week and Space Technology,” McGraw-Hill, New York.
- “Avionics Magazine,” Phillips Business Information, Inc., Potomac, MD.



Aircraft Speed and Altitude

William Gracey (deceased)

NASA-Langley Research Center

Thomas M. Weeks

Ball Aerospace & Technologies Corp.

Tony Whitmore

NASA Dryden Flight Research Center

- I. Instrument Systems and Errors
- II. Equations for Aircraft Speed and Altitude
- III. Pressure Sensors
- IV. Pressure Installations
- V. Flight Calibration Methods
- VI. Air Data Measurements in the Hypersonic Flight Environment
- VII. Altimetry

GLOSSARY

Ambient pressure Local barometric pressure at the flight altitude.

Calibrated airspeed Indicated airspeed corrected for position error; equal to true airspeed at sea level in the standard atmosphere.

Freestream static pressure Local pressure sensed parallel to the direction of flow outside of the influence of the aircraft. For a flying vehicle, the ambient and freestream static pressure levels can be taken to be essentially identical. For a wind tunnel test, this is generally not an accurate assumption.

Freestream total (stagnation) pressure Generally, a quantity defined as the pressure level that would be experienced if the freestream flow conditions were slowed from the current velocity to zero velocity.

Geometric height Height above sea level or above the terrain.

Ground speed Speed of an aircraft relative to the ground.

Hypersonic flow conditions For high flight Mach numbers, nominally above 6, a very significant temperature rise occurs across the bow shock wave and the fluid no longer behaves as a calorically perfect gas. The large temperature rise across the shock wave causes vibration and modes of the diatomic air molecules to ex-

cite, which draws energy away from the flow. Thus, the stagnation temperatures actually encountered are considerably lower than those which would be computed if perfect gas calculations (with $\gamma = 1.4$) were used. Still higher Mach numbers lead to dissociation of the air molecules. Under such conditions, aerodynamics gives way to aerothermodynamics. Airspeed considerations in this regime of flight are considered outside the scope of this article.

Impact pressure Difference between the local airstream total (or stagnation) pressure and the local airstream static pressure.

Indicated airspeed Speed indicated by airspeed indicator (uncorrected for position error).

Indicated altitude Pressure altitude indicated by altimeter (uncorrected for position error).

Mach number Ratio of true airspeed to the speed of sound.

Position error Difference between pressure measured by static-pressure installation and free-stream static pressure.

Pressure altitude Height corresponding to a given pressure in the standard atmosphere; indicated altitude corrected for position error.

Standard atmosphere Pressure-height relation based on average atmospheric conditions at latitude 45°N.

Static pressure Local pressure sensed parallel to the direction of flow at any given point in the airstream about an aircraft.

Subsonic flow conditions Most commercial and general aviation aircraft cruise at speeds below the ambient speed of sound at the cruise altitude. As long as the flight speed remains less than about 80% of the speed of sound (Mach 0.8), the local velocity around the aircraft will remain less than sonic and the flowfield itself may be described as "subsonic."

Supersonic flow conditions Increases in flight speed will eventually result in local flow speeds in all regions of the flowfield surrounding the aircraft to exceed the local speed of sound. This generally happens when the flight speed has reached about 120% of the speed of sound at the flight altitude (Mach 1.2). At this flight condition, bow shocks form at the bow of the aircraft and along the leading edges of wings and tails. Shocks also form at trailing edges.

Total pressure (stagnation pressure) Generally, a quantity defined as the pressure level that would be experienced at a point in the airstream about an aircraft if the flow conditions at that point were slowed from the current velocity to zero velocity.

Transonic flow conditions As cruise speeds increase beyond 80% of the speed of sound at the flight altitude, some portions of the flowfield at points about

the aircraft will exceed the local speed of sound at those points. The appearance of local pockets of supersonic flow, accompanied by shock waves, signals the onset of "transonic" flow conditions. This condition may be observed in commercial aircraft flying at altitudes of approximately 35,000 feet and, of course, high-performance military aircraft.

True airspeed Speed of an aircraft relative to the surrounding air mass.

True dynamic pressure Difference between the free-stream total or stagnation pressure and the freestream static pressure.

SPEED AND ALTITUDE of an aircraft are measured by an airspeed indicator, a true-airspeed indicator, a Mach meter, a rate-of-climb (or vertical speed) indicator, and an altimeter (Gracey, 1981). Measurements of a high order of accuracy are required for the safe and efficient operation of an aircraft. Accurate measurements of speed, for example, are needed to avoid loss of control at low speeds (stall condition) and to avoid exceeding the aerodynamic and structural limitations of an aircraft at high speeds. Accurate measurements of altitude are needed to ensure clearance of terrain obstacles and to maintain vertical separation minima along airways.

All of the speed and altitude instruments are actuated by pressures, while one, the true-airspeed indicator, is also actuated by air temperature. The pressures used to actuate the instruments are the total and static pressure. The static pressure is the atmospheric pressure at the flight level of the aircraft, and the total pressure is the sum of the static pressure and the impact pressure, which is the pressure developed by the forward speed of the aircraft.

The pressures are sensed by static-pressure and total-pressure tubes or by static ports located on the sides of the fuselage. For most installations, the measurement of the total pressure is essentially free of error. The measurement of the static pressure, however, can differ considerably from the atmospheric pressure at the flight level because of the design of the sensor or its location on the aircraft. For accurate measurements of speed and altitude, therefore, it is essential that the errors of the static-pressure system be determined and that corrections for these errors be applied.

I. INSTRUMENT SYSTEMS AND ERRORS

Diagrams of two instrument systems for the measurement of speed and altitude are shown in Fig. 1. All of the instruments are connected to the static-pressure tube, which has orifices either on the body of the tube or mounted flush to the aircraft surface. The three forward-speed in-

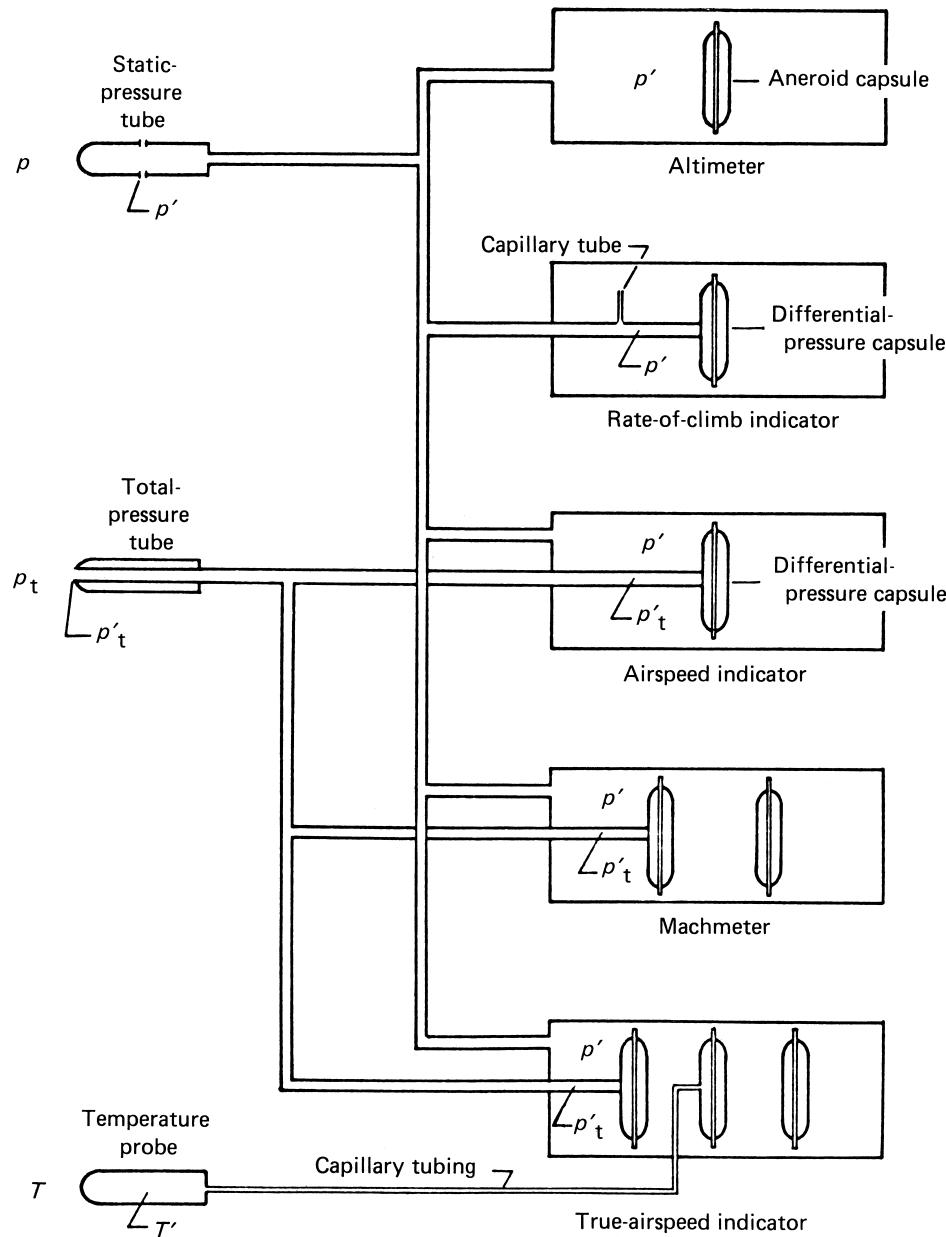


FIGURE 1 Instrument system for the measurement of speed and altitude. [From Gracey, W. (1981). "Measurement of Aircraft Speed and Altitude," Wiley-Interscience, New York.]

struments are connected to an open-ended tube called a total-pressure (or pitot) tube. Since the total pressure is the sum of the static pressure and the impact pressure, the impact pressure q_c can be expressed by:

$$q_c = p_t - p \quad (1)$$

where p and p_t are the free-stream static and total pressures, respectively. The pressures sensed by the tubes, however, and the temperature sensed by the temperature probe can differ from the free-stream values. These dif-

ferences, called errors, are expressed by:

$$\Delta p = p' - p \quad (2)$$

$$\Delta p_t = p'_t - p_t \quad (3)$$

$$\Delta T = T' - T \quad (4)$$

where p' , p_t and T' are the measured pressures and temperature.

Within the instrument cases, the pressures and temperature are measured with two types of capsule: an

aneroid capsule that is evacuated and sealed and a differential-pressure capsule that has an opening connected to a pressure source. The aneroid capsule measures the pressure inside the case, while the differential-pressure capsule measures the difference between the pressure in the capsule and the pressure in the case. The pressure-sensing element in the altimeter is an aneroid capsule that expands as the static pressure inside the case decreases with increasing altitude.

In the case of the “rate-of-climb indicator,” the static pressure tube is connected to a differential-pressure capsule and to a capillary tube that opens into the instrument case. With a change in static pressure, the simultaneous flow of air into, or out of, the capsule and the capillary tube is adjusted (by the size of the capillary leak) so that the capsule deflects in terms of a rate of change of pressure, which is calibrated to yield a measure of vertical speed.

The pressure-sensing element of the airspeed indicator is a differential-pressure capsule that expands as the total pressure increases. Since the pressure inside the case is the static pressure, the instrument performs a mechanical subtraction of the total and static pressures to yield a measure of the impact pressure in accordance with Eq. (1).

The Machmeter contains both an aneroid capsule and a differential-pressure capsule to provide measures of static pressure and impact pressure. The deflections of the two capsules are coupled to yield, mechanically, the ratio of impact pressure to static pressure (q_c/P), which is a function of Mach number.

The true-airspeed indicator contains: (1) two differential-pressure capsules to provide measures of impact pressure and air temperature, and (2) an aneroid capsule to provide a measure of static pressure. Since the true airspeed is a function of the dynamic pressure (derived from the measured impact and static pressures, as discussed in Section II,D) and the air density (derived from the static pressure and temperature), the deflections of the three capsules can be coupled to yield a measure of true airspeed.

The scales on the dials of the instruments that measure speed and altitude are calibrated in accordance with equations that relate the measured pressures and temperature with the flight parameters.

II. EQUATIONS FOR AIRCRAFT SPEED AND ALTITUDE

A. Altitude Equations

Although the static pressure at any given point in the atmosphere varies from day to day, the decrease in the static pressure with height is generally continuous at any one

time and place. Accordingly, a pressure–height relation based on average atmospheric conditions has been adopted as a standard for the measurement of static pressure. This so-called standard atmosphere is based on mean values of pressure, temperature, density, and the acceleration due to gravity at sea level and on a mean value of the variation of temperature with height, all at a latitude of 45°N.

The basic equation for the variation of pressure with height is

$$dp = -g\rho dZ \quad (5)$$

where p is the pressure, g the acceleration due to gravity, ρ the air density, and Z the geometric height. For the standard atmospheres developed from 1922 to 1935, the height parameter was the geometric height. For later atmospheres (1955–1962), the decrease of g with height was taken into account with the formation of a new height parameter called geopotential altitude, H . The relation between H and Z is given by:

$$dH = \frac{g}{g_0} dZ \quad \text{or} \quad H = \int_0^z \frac{g}{g_0} dZ \quad (6)$$

where the subscript 0 refers to sea level. The value of Z/H varies uniformly from 1.0 at sea level to 1.0048 at 100,000 ft.

The relation between pressure and geopotential altitude is derived from Eq. (5) (with the height parameter in terms of H) and the perfect gas law expressed as:

$$\rho = p W_m / g R^* T \quad (7)$$

where W_m is the mean molecular weight of air, R^* the universal gas constant, and T the temperature. For the atmospheric layers in which the temperature varies linearly with height ($H < 36,090$ ft and $H > 65,800$ ft), the relation between p and H is given by:

$$\begin{aligned} \ln p &= \ln p_b - \frac{g_0 W_{m_0}}{(dT/dH)R^*} \\ &\times \ln \frac{(dT/dH)(H - H_b) + T_b}{T_b} \end{aligned} \quad (8)$$

and for the constant temperature layer between 36,090 and 65,800 ft, the relationship is given by:

$$\ln p = \ln p_b - \frac{g_0 W_{m_0}}{T_b R^*} (H - H_b) \quad (9)$$

where dT/dH is the temperature gradient and the subscript b is a base or reference level.

Pressure-altitude tables derived from Eqs. (8) and (9) are used for the calibration of altimeters and rate-of-climb indicators. Tables of static pressure versus altitude in both U.S. Customary Units and the International System of Units (SI) are given in [Gracey \(1981\)](#).

B. Airspeed Equations

In incompressible flow, the pressure developed by the forward motion of a body is called the dynamic pressure q , which is related to the true airspeed V by:

$$q = \frac{1}{2} \rho V^2 \quad (10)$$

where ρ is the density of the air and V the speed of the body relative to the air. Air, however, is compressible, and when airspeed is measured with a pitot-static tube, the air is compressed as it is brought to a stop in the pitot tube. As a consequence of this compression, the measured impact pressure is higher than the dynamic pressure. The effects of compressibility are taken into account in the following equations, in which the calibrated airspeed V_c is made equal to the true airspeed V at standard sea-level conditions:

$$q_c = p_0 \left[\left(1 + \frac{\gamma - 1}{2\gamma} \frac{\rho_0}{p_0} V_c^2 \right)^{\gamma/(\gamma-1)} - 1 \right] \\ V_c \leq a_0 \quad (11)$$

$$q_c = \frac{1 + \gamma}{2} \left(\frac{V_c}{a_0} \right)^2 p_0 \\ \times \left[\frac{(\gamma + 1)^2}{4\gamma - 2(\gamma - 1)(a_0/V_c)^2} \right]^{1/(\gamma-1)} - p_0 \\ V_c \geq a_0 \quad (12)$$

where a is the speed of sound, γ is the ratio of the specific heats of air (with a value of 1.4), and the subscript 0 refers to values at sea level. Airspeed indicators are calibrated in accordance with Eq. (11) for subsonic speeds and Eq. (12) for supersonic speeds. Tables of impact pressure versus calibrated airspeed in both U.S. Customary Units and the International System of Units are given in [Gracey \(1981\)](#).

Since the airspeed indicator measures true airspeed only at standard sea-level conditions, the calibrated airspeeds at altitude will be lower than the true airspeeds, as shown by the following equation,

$$V = V_c \frac{f}{f_0} \sqrt{\frac{\rho_0}{\rho}} \quad (13)$$

where ρ is the density and f a compressibility factor defined by:

$$f = \sqrt{\frac{\gamma}{\gamma - 1} \frac{p}{q_c} \left[\left(\frac{q_c}{p} + 1 \right)^{(\gamma-1)/\gamma} - 1 \right]} \quad (14)$$

C. Mach-Number Equations

The Mach number M is defined as the ratio of the true airspeed V to the speed of sound a in air at the flight altitude:

$$M = V/a \quad (15)$$

For the calibration of Machmeters, the Mach number can be related to the ratio of the impact pressure to the static pressure q_c/P by the following expressions:

$$\frac{q_c}{p} = \left(1 + \frac{\gamma - 1}{2} M^2 \right)^{\gamma/(\gamma-1)} - 1, \quad M \leq 1 \quad (16)$$

$$\frac{q_c}{p} = \frac{1 + \gamma}{2} M^2 \left[\frac{(1 + \gamma^2)/M^2}{4\gamma M^2 - 2(\gamma - 1)} \right]^{1/(\gamma-1)} - 1. \quad (17)$$

$$M \geq 1$$

Machmeters are calibrated in accordance with Eq. (16) for subsonic speeds and Eq. (17) for supersonic speeds. Tables of impact pressure/static pressure versus Mach number can be found in [Gracey](#).¹

D. True Airspeed Equations

From Eq. (10), the true airspeed V is seen to be a function of the dynamic pressure q and the air density ρ as follows:

$$V = \sqrt{2q/\rho} \quad (18)$$

As shown by the following equations, the dynamic pressure is a function of the static pressure p and the Mach number M :

$$q = \gamma p M^2 / 2 \quad (19)$$

and the air density ρ is a function of the static pressure and the air temperature T :

$$\rho = \rho_0 \frac{p T_0}{p_0 T} \quad (20)$$

Since Mach number is derived from measurements of q_c and p , the measurement of these pressures can be combined with the measurement of T to provide indications of true airspeed, as discussed earlier. In contrast to the airspeed indicator, which indicates true airspeed only at sea level (under standard conditions), the true-airspeed indicator indicates true airspeed independent of altitude. [Table I](#) gives values of true airspeed V in knots for values of calibrated airspeed V_c in knots and pressure altitude H in geopotential feet.

The airspeed indicator, true-airspeed indicator, and Machmeter measure speed with respect to the air mass. Since the air mass can move with respect to the ground, the measurement of ground speed, the speed of basic importance to air navigation, must be derived from ground navigational aids or an inertial guidance system.

TABLE I True Airspeed V for Values of Calibrated Airspeed V_c and Values of Pressure Altitude H^a

H (geopotential ft)	V_c (knots)									
	100	200	300	400	500	600	700	800	900	1000
0	100.0	200.0	300.0	400.0	500.0	600.0	700.0	800.0	900.0	1000
5000	107.7	215.0	321.6	427.4	532.2	635.8	740.3	847.3	955.2	1064
10,000	116.2	231.6	345.4	457.2	566.8	674.5	785.0	900.5	1018	1136
15,000	125.8	250.0	371.5	489.4	603.8	716.3	835.2	960.9	1089	1218
20,000	137.2	270.5	400.1	524.4	643.4	763.0	892.4	1030	1170	1310
25,000	148.7	293.4	431.5	562.0	686.6	816.2	958.0	1109	1263	1418
30,000	162.4	318.9	465.9	602.6	735.4	877.5	1034	1201	1370	1541
35,000	178.0	347.4	503.5	646.9	791.6	948.7	1122	1307	1494	1682
40,000	199.1	385.6	553.7	708.9	871.5	1049	1245	1454	1666	1878
45,000	223.7	429.1	610.0	782.4	967.0	1169	1392	1629	1869	
50,000	251.0	476.4	671.6	865.7	1076	1306	1559	1827		
55,000	281.3	527.3	740.3	960.3	1199	1460	1747			
60,000	314.9	581.8	817.9	1068	1340	1636	1961			
65,000	351.8	640.4	906.0	1191	1499	1835				
70,000	394.8	709.9	1013	1338	1690	2073				
75,000	440.3	785.3	1130	1501	1901					
80,000	489.4	870.3	1263	1684	2139					
85,000	540.2	962.9	1408	1885						
90,000	596.2	1071	1576	2111						
95,000	656.2	1193	1766							
100,000	722.2	1330	1979							

^a True-airspeed values V are given in knots.

From Gracey, W. (1981). "Measurement of Aircraft Speed and Altitude," Wiley-Interscience, New York.

III. PRESSURE SENSORS

For the instrument system shown in Fig. 1a, the static pressure and total pressure are sensed by separate tubes. Examples of a total-pressure (or pitot) tube and a static-pressure tube are shown in Fig. 2. The two tubes can be combined into a single tube called a *pitot-static tube*. The static pressure can also be sensed by orifices on the side of the fuselage (called *fuselage vents* or *static ports*); in this case, a pitot tube is used in combination with the vents to form another type of pitot-static installation.

A. Total-Pressure Tubes

The installation of a pitot tube is relatively simple with respect to both the design of the tube and its location on the aircraft. At subsonic speeds, almost any open-ended tube will measure total pressure correctly when the tube is aligned with the airflow. When the tube is inclined to the flow, however, the measured pressure begins to decrease at some angle of inclination. With early pitot tube designs, in which the nose shape is hemispherical and the impact opening a small percentage of the frontal area of the tube, the total pressure begins to decrease at only a few de-

grees of inclination. For later designs, in which the impact opening is as large as the frontal area (e.g., the pitot tube in Fig. 2), the total pressure does not begin to decrease until an angle of inclination of $\sim 25^\circ$ has been reached.

Locations of total-pressure tubes that have proved satisfactory for subsonic operations include positions ahead of the fuselage nose, the wing, and the vertical fin for tubes mounted on short horizontal booms and positions along the fuselage or under the wing for tubes mounted on short struts (see Fig. 2). For operations at supersonic speeds, the tube should be located ahead of shocks emanating from any part of the aircraft. The position that best meets this requirement is obviously a position ahead of the fuselage nose. In this case, the tube is ahead of the fuselage bow shock and is influenced only by the small normal shock that forms ahead of the tube. The total-pressure loss through this shock is taken into account in the airspeed and Mach number equations for supersonic speeds [Eqs. (12) and (17)].

B. Static-Pressure Tubes

The pressures sensed by the orifices of a static-pressure tube depend primarily on the location of the orifices aft of

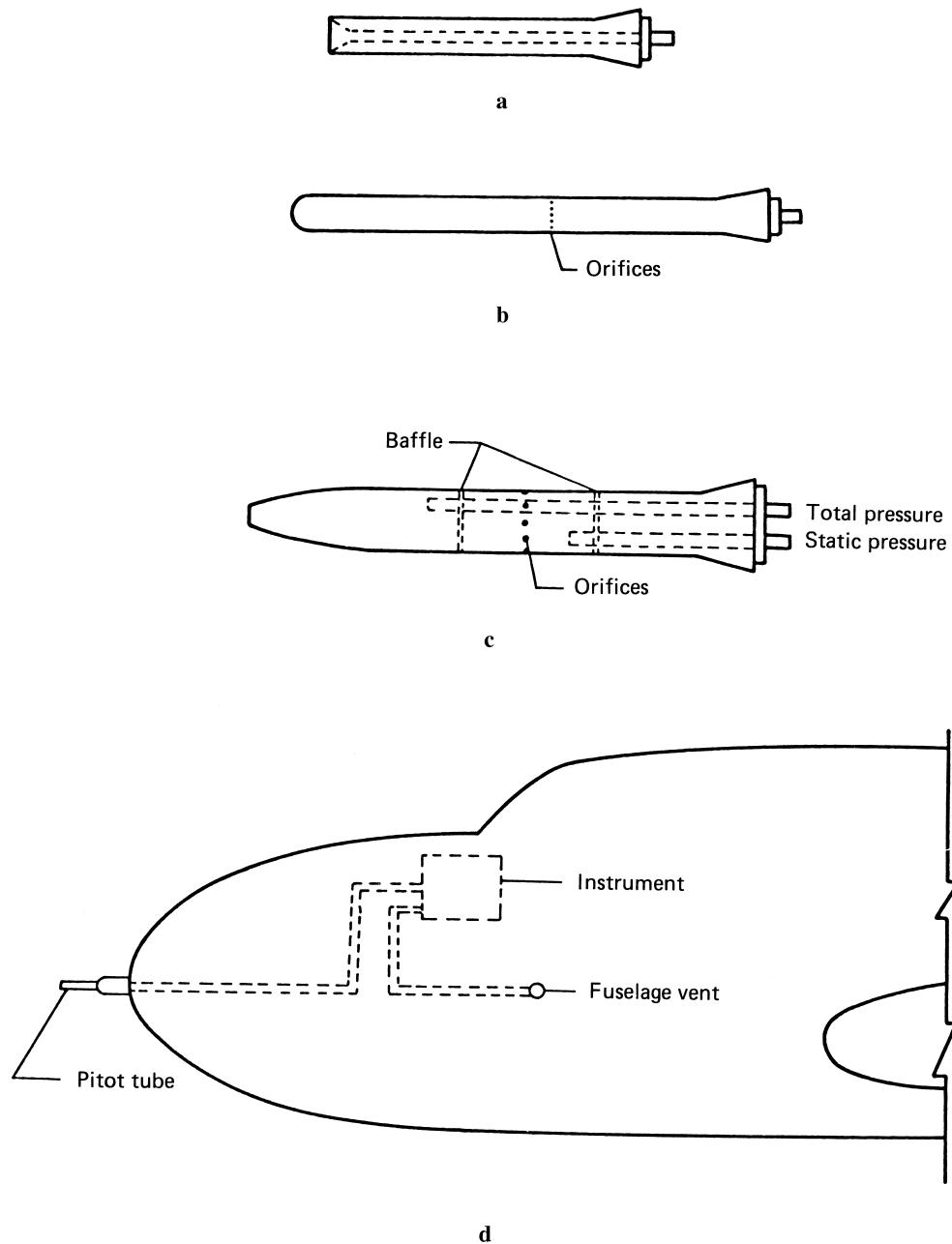


FIGURE 2 Three pressure tubes—(a) pitot, (b) static-pressure—and (c) pitot-static, and a pitot tube/fuselage vent installation (d). [From Gracey, W. (1981). "Measurement of Aircraft Speed and Altitude," Wiley-Interscience, New York.]

the nose of the tube and on the size and location of collars or struts to the rear of the orifices. The pressures near the nose of the tube are below the free-stream pressure; with increasing distance from the nose, the pressures approach the free-stream pressure and reach that value at about five tube diameters in subsonic flow and at about eight tube diameters in supersonic flow. The pressures ahead of a collar or strut, on the other hand, are above the free-stream pressure. Thus, the problem of designing a static-

pressure tube is one of balancing the lower-than-free-stream pressures due to the location of the orifices behind the nose and the higher-than-free-stream pressures due to the location of the orifices ahead of the collar or strut. The effect of a strut or other body in creating a positive flow field upstream from the body is called the *blocking effect*.

An example of a pitot-static tube in which the positive and negative pressure errors at the orifices have been

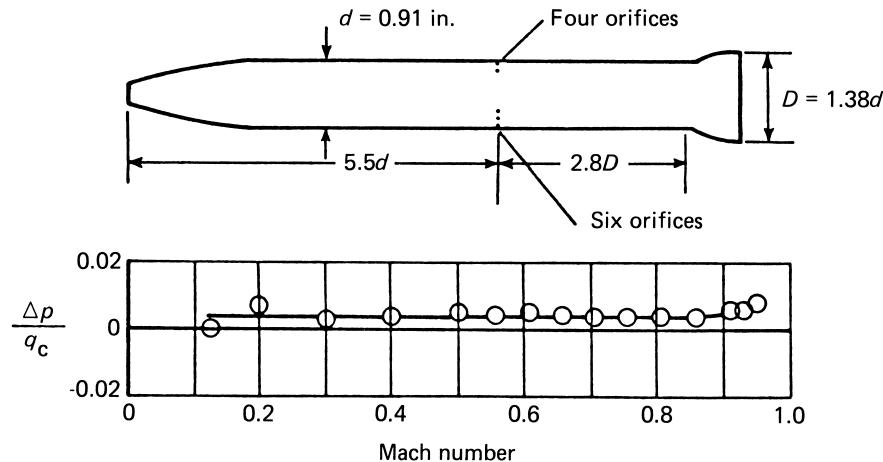


FIGURE 3 Calibration of a pitot–static tube aligned with the flow. [From Gracey, W. (1981). “Measurement of Aircraft Speed and Altitude,” Wiley–Interscience, New York.]

very nearly balanced is shown in Fig. 3. As shown by the calibration of this tube (when aligned with the flow), the static-pressure error Δp as a fraction of the impact pressure q_c is $\sim 0.5\%$ throughout the subsonic speed range. The orifices of this tube are spaced around the top and bottom of the tube to increase the range of angle of attack (inclination in a vertical plane) through which the measured pressures remain unaffected. For this orifice configuration, the measured pressures remain within $\sim 0.5\% q_c$ through an angle of attack range of $\sim 20^\circ$. In contrast, for tubes in which the orifices encircle the tube, the measured angle of attack range through which the pressures remain within $\sim 0.5\% q_c$ is less than 5° .

IV. PRESSURE INSTALLATIONS

For a steady flow condition, the flow of the air over a body creates a pressure field in which the static pressures vary from point to point, while the total pressure at all points remains the same. It is for this reason that the measurement of local static pressure on an aircraft is more complicated than the measurement of total pressure. The pressure field created by the airflow may change with the configuration of the aircraft and with Mach number and angle of attack. Thus, for a given aircraft configuration, the problem of designing a static-pressure-measuring system is primarily one of finding a location where the static-pressure error varies by the least amount throughout the operating range of the aircraft.

A. Flow Field

The variation of the pressures in the flow field can be described by Bernoulli’s equation for the total pressure P_t in incompressible flow:

$$p_t = p_1 + \frac{1}{2}\rho V_1^2 = \text{const} \quad (21)$$

where p_t is the local static pressure and V_1 the local flow velocity. As expressed by Eq. (21), the total pressure in incompressible flow remains constant (at the free-stream value) at all points in the flow field, whereas the local static pressure at any point varies inversely as the square of the local velocity.

The variation of the local static pressure with the local velocity is illustrated by a diagram of the flow around a fuselage-like body in Fig. 4. The five lines of flow (streamlines) shown in this figure represent the paths of the individual particles of air. At a great distance ahead of the body and at a great distance behind, the streamlines are parallel and the total pressure, static pressure, and velocity of the particles on each of the streamlines are the free-stream values.

Relative magnitudes of the local pressure and local velocity at three points near the nose of the body are shown in Fig. 4. At a position directly ahead of the nose, the local static pressure is higher than the free-stream static pressure, while at a position just aft of the nose, the local static pressure is lower than the stream value. At a point on the leading edge of the nose, where the air particles come to a stop, the local static pressure is equal to the free-stream total pressure.

The flow field in Fig. 4 applies to incompressible flow or to compressible flow at very low speeds. For higher speeds in compressible flow, the flow field changes markedly, particularly at transonic (corresponding to a flight Mach number between 0.8 and 1.2) and supersonic speeds. In the subsonic speed range, the flow field extends in all directions from the aircraft. The difference between the local static pressure and the free-stream static pressure is greatest in the vicinity of the aircraft and decreases with distance from it. In the transonic speed range, the flow field is altered by shock waves that form along the lines of maximum curvature of the fuselage, wings, and tail

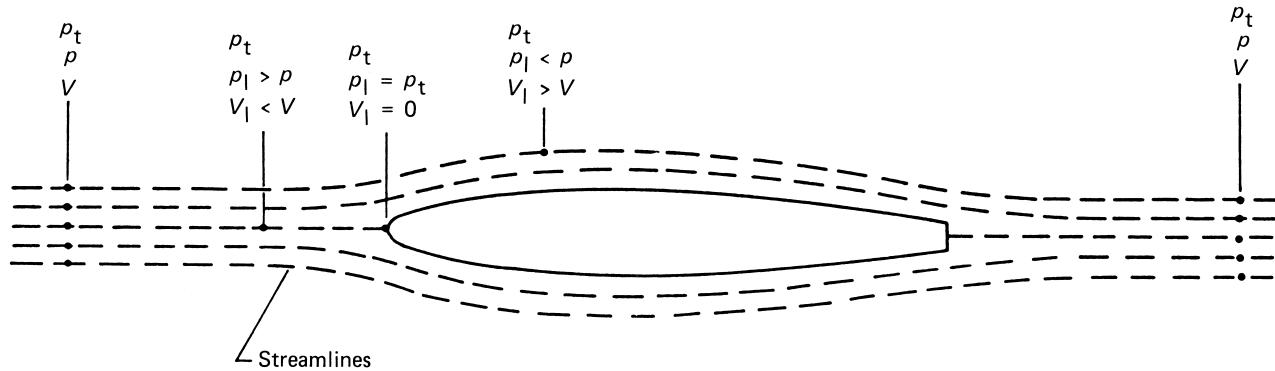


FIGURE 4 Local pressures and velocities in vicinity of a fuselage-like body [see Eq. (21)]. [From Gracey, W. (1981). "Measurement of Aircraft Speed and Altitude," Wiley-Interscience, New York.]

surfaces. At supersonic speeds, the flow field is confined to the regions behind the shock wave that forms ahead of the fuselage nose (fuselage bow shock).

B. Pitot-Static Installations

Pitot-static tubes with strut supports (Fig. 5) are generally attached to the underside of a wing or, in pairs, to the sides of the fuselage. Tubes designed for end mounting on horizontal booms are installed on the nose of the fuselage, the outboard section of the wing, or the tip of the verti-

cal fin. Figure 6 shows three boom-mounted pitot-static installations and a fuselage vent system.

For the fuselage vent system, the measured static pressure p' is usually the same as the local static pressure p_1 . With the pitot-static tube installations, on the other hand, the local static pressure is altered by the presence of the tube, because the tube creates a small flow field of its own. For all of the installations, the static-pressure error Δp of the installation (also called *position error*) is the difference between the measured pressure and the free-stream pressure:

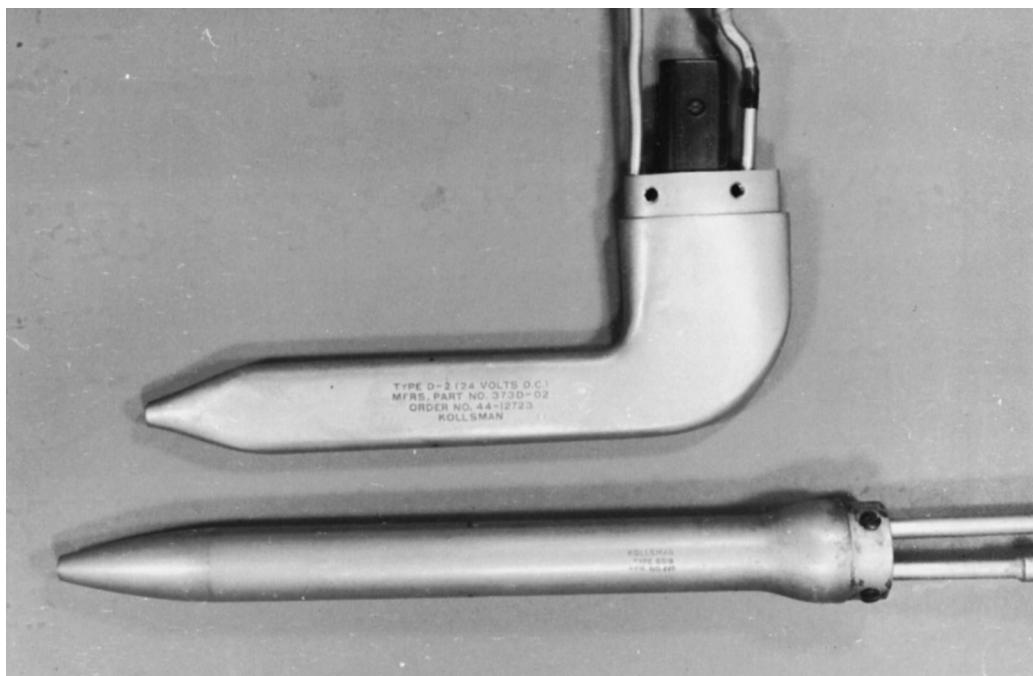


FIGURE 5 Pitot-static tubes designed for strut mounting and end mounting. [From Gracey, W. (1981). "Measurement of Aircraft Speed and Altitude," Wiley-Interscience, New York.]

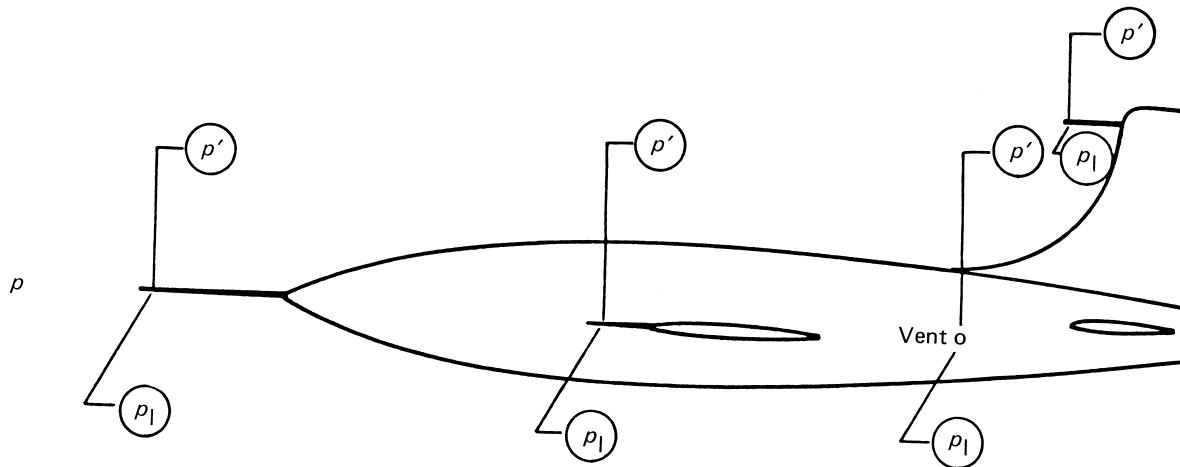


FIGURE 6 Three boom-mounted pitot–static installations and a fuselage vent system. See text for definition of variables. [From Gracey, W. (1981). “Measurement of Aircraft Speed and Altitude,” Wiley–Interscience, New York.]

$$\Delta p = p' - p \quad (22)$$

For tubes installed on booms ahead of the fuselage, wing, or vertical fin, the position error decreases as the length of the boom is increased. For fuselage vent installations, the local pressure at the vent is determined by the position of the vent along the fuselage and by its position ahead of any part of the aircraft (such as the wing or the tail surfaces) that creates a positive pressure error at the vent. Thus, the problem of designing a fuselage vent system is similar to that for static-pressure tubes, that is, balancing the positive and negative pressure errors at the vents.

The position errors of static-pressure installations are determined by flight calibration of the installation. The calibrations for low subsonic speeds (up to a Mach number of ~ 0.5) are conducted at low altitudes, whereas the calibrations for high subsonic and supersonic speeds are conducted at high altitudes.

C. Nose-Boom and Wing-Tip Installations

The calibrations of a nose-boom installation and a wing-tip installation are shown in Fig. 7. With the nose-boom installation, the position error in the mid-subsonic range is positive due to the blocking effect of the fuselage nose. As the speed is decreased to the stall speed, the error decreases due to the effects of lift coefficient (a function of angle of attack). In the high subsonic range, the error increases rapidly due to the increases in the effects of compressibility with Mach number, reaches a peak value at a Mach number just beyond 1.0, then falls abruptly as the fuselage bow wave passes over the orifices on the tube. At this point, the orifices are isolated from the flow field of the aircraft, so that the position error of the installation becomes that of the static-pressure tube.

With the wing-tip installation, the error in the mid-subsonic range is positive but smaller than that for the nose-boom installation because of the smaller blocking effect of the wing. At the lower speeds, the error curve moves down as in the case of the nose-boom installation, but in this case the variation is greater because of the larger effects of lift coefficient. In the upper subsonic range, the error increases to a peak value, falls to a negative value when the shock wave ahead of the wing passes the orifices, then rises to positive values as the orifices come under the influence of the fuselage bow shock.

An important point about the calibrations in Fig. 7 is the fact that the variations of the position errors for these installations are characteristic of nose-boom and wing-tip installations. For example, if the boom length for the nose-boom installation were shorter, the entire curve would move upward. The decrease in the error near the stall would be somewhat greater, and the increase in the error in the upper Mach range would be greater, but the general shape of the curve would be similar to that shown in Fig. 7.

D. Fuselage Vent Installations

In contrast to the position errors of the noseboom and wing-tip installations in the mid-subsonic range, which are in all cases positive, the position errors of fuselage vent installations can be either positive or negative, depending on the location of the vents along the fuselage and on their position with respect to the wing or tail surfaces. Figure 8 shows the calibrations of fuselage vent installations on two types of transport aircraft. For one, the position error is positive in the mid-subsonic range and rises in the upper subsonic range due to the increase in the blocking effect of the wing. For the other, the errors are negative at both

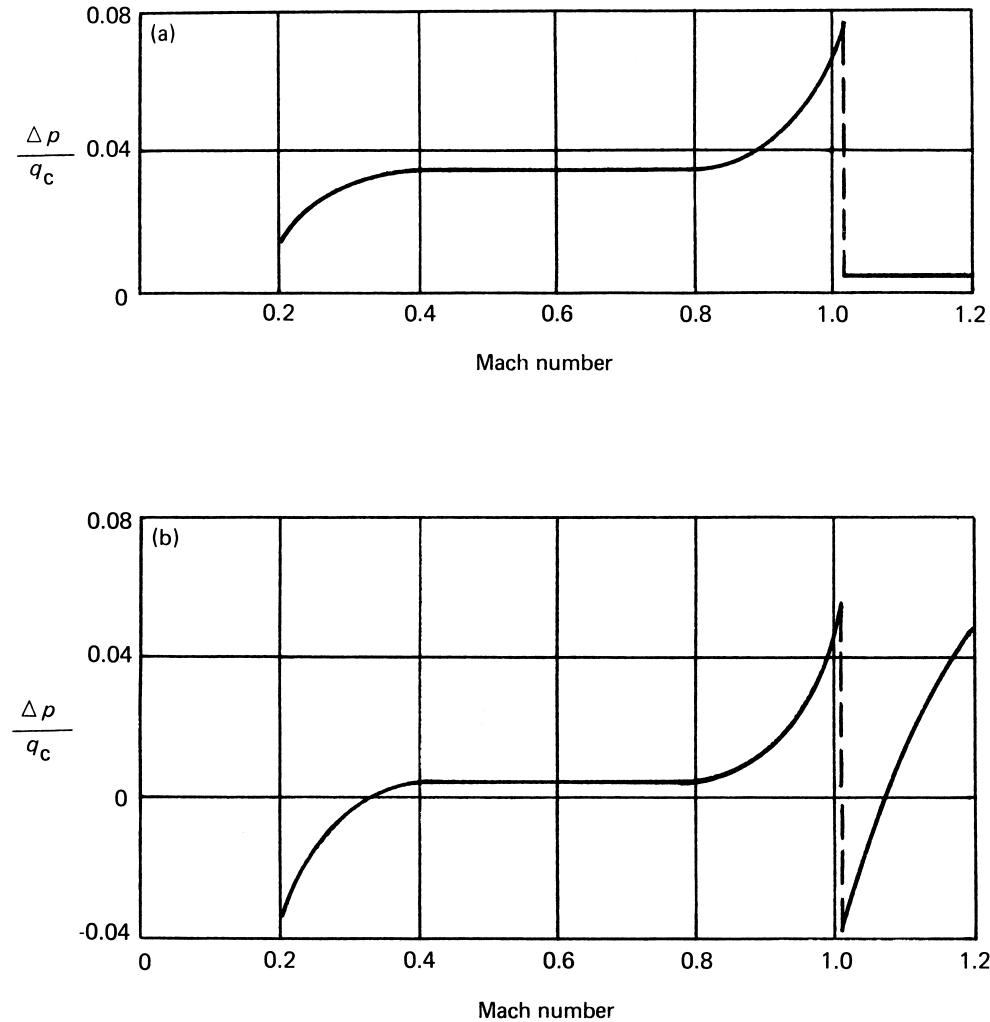


FIGURE 7 Calibration of (a) nose-boom and (b) wing-tip installations on two types of experimental aircraft.

mid-subsonic and upper subsonic speeds; at Mach numbers above 0.8, the calibration curve moves upward, as in the case of the calibration of the first transport.

For flight operations, the position errors are converted to errors in airspeed and altitude so that corrections can be applied to the values of the indicated airspeed and indicated altitude. Figure 9 shows the altitude errors ΔH and the airspeed errors ΔV_c that correspond to position errors of 1% of the impact pressure ($\Delta p/q_c = 0.01$). The corrections for the two errors can be applied either manually by the use of correction charts or automatically by means of air data computers.

V. FLIGHT CALIBRATION METHODS

For the determination of position errors of a static-pressure installation, the basic problem is that of determining the

free-stream static pressure p at the flight level of the aircraft. The position error Δp is then determined as the difference between the pressure p' measured by the installation and the free-stream pressure; that is, $\Delta p = p' - p$.

Figure 10 shows five of the most common methods for determining the free-stream static pressure at the flight altitude. With the tower method, the aircraft is flown past the top of a tall tower, where the free-stream static pressure is measured with a barometer or an altimeter. The height of the aircraft with respect to the top of the tower is determined with a camera located at the top of the tower, and the pressure at the height of the aircraft is determined from a pressure-height survey based on measurements of the pressures at various points along the height of the tower.

The ground-based radar method is used for calibrations at high altitudes. This method involves a two-step procedure: a pressure-height survey and a series of high-speed test runs. For the survey, the aircraft installation is first

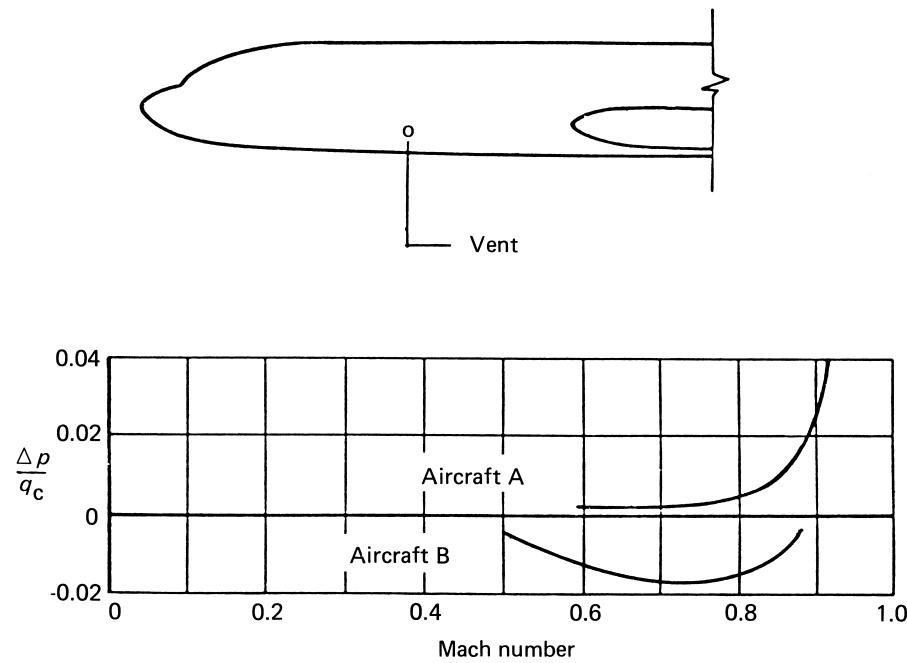


FIGURE 8 Calibrations of fuselage vent installations on two type of transport aircraft.

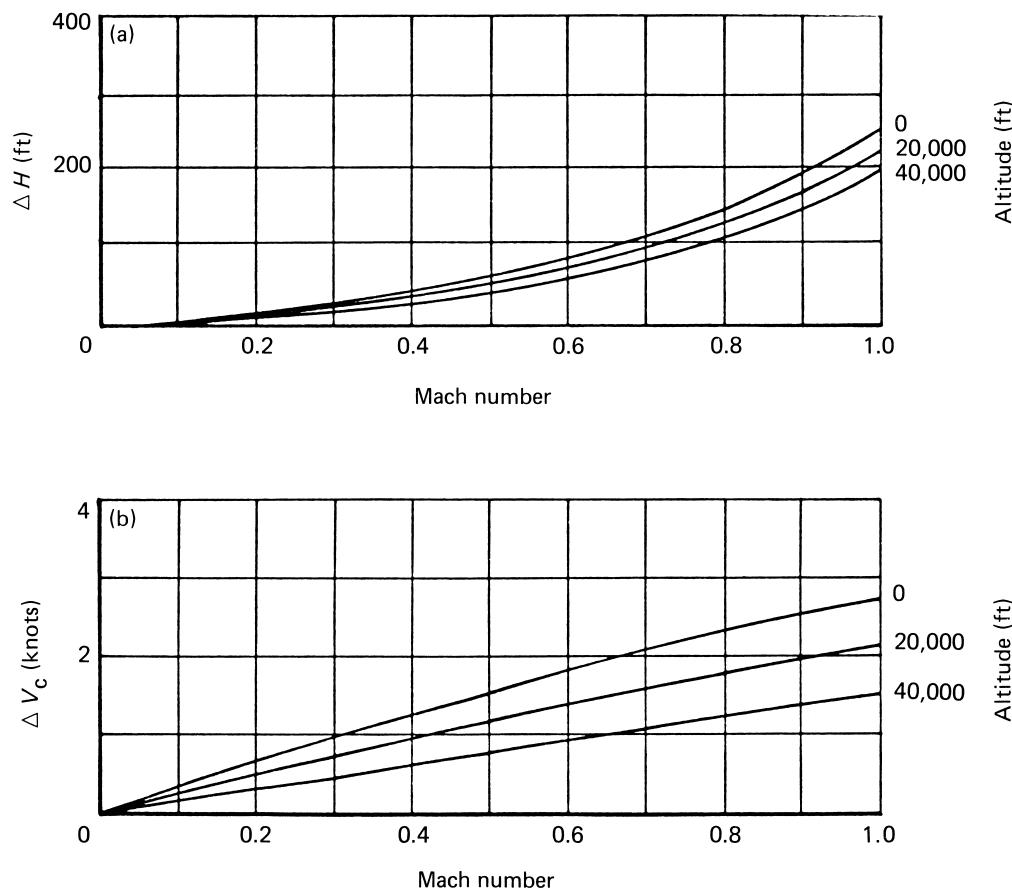


FIGURE 9 Altitude errors (ΔH) (a) and airspeed errors ΔV_c (b) corresponding to a static-pressure error of 1% of the impact pressure. [From Gracey, W. (1981). "Measurement of Aircraft Speed and Altitude," Wiley-Interscience, New York.]

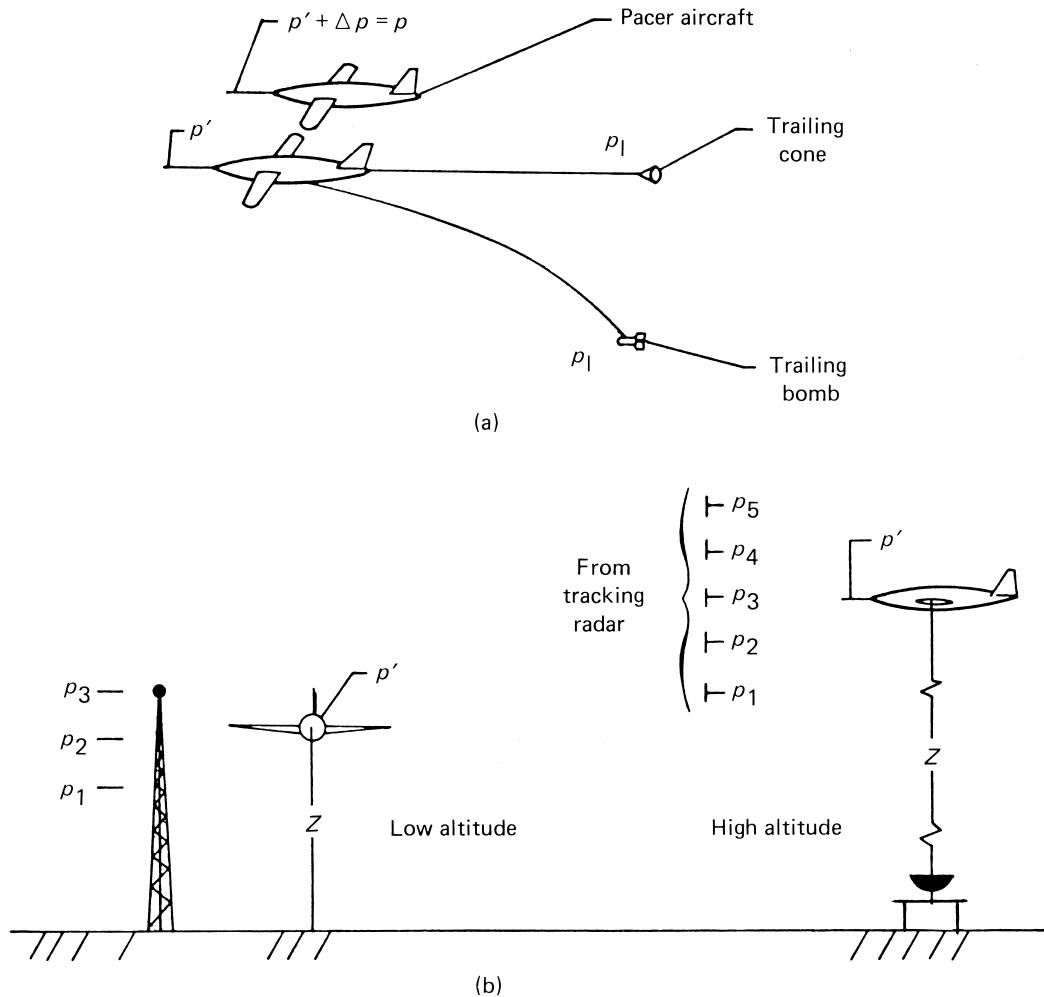


FIGURE 10 Five methods for determining free-stream static pressure p at the flight level of the aircraft. (a) Static pressure p measurement at reference pressure and source below, behind, or alongside aircraft; (b) p derived from measurement of height of aircraft and pressure gradient at test altitude range. [From Gracey, W. (1981). "Measurement of Aircraft Speed and Altitude," Wiley-Interscience, New York.]

calibrated at a low airspeed by a low-altitude method such as the tower method. The aircraft is then flown at this low airspeed at a series of altitudes. At each altitude, the height of the aircraft is measured by the radar, and the free-stream static pressure at that height is determined from the pressure p' measured by the installation and the position error Δp of the installation at the low airspeed. In the second step, the aircraft is flown through the test atmosphere in a series of high-speed runs. For each test run, the height of the aircraft is measured by the radar and the free-stream static pressure at that height is determined from the pressure-height survey.

With each of the three methods shown at the top of Fig. 10, the free-stream static pressure is measured with a reference pressure source that is moving with the aircraft. With the pacer aircraft method, the reference source is the

calibrated installation of another aircraft flying in formation with the test aircraft. The installation of the reference aircraft will have been calibrated by one or more of the other calibration methods.

With the trailing cone method, the reference pressure source is a set of orifices located ahead of the cone in the tubing connecting the cone to a differential-pressure instrument in the aircraft. For this method, the cone is trailed at a sufficient distance behind the aircraft where the local static pressure is very nearly equal to the free-stream static pressure.

With the trailing bomb method, the reference pressure source is a set of orifices on the body of a bomb (Fig. 11), which is attached to tubing that is connected to a differential-pressure instrument in the aircraft. In this case, the bomb is trailed at a sufficient distance

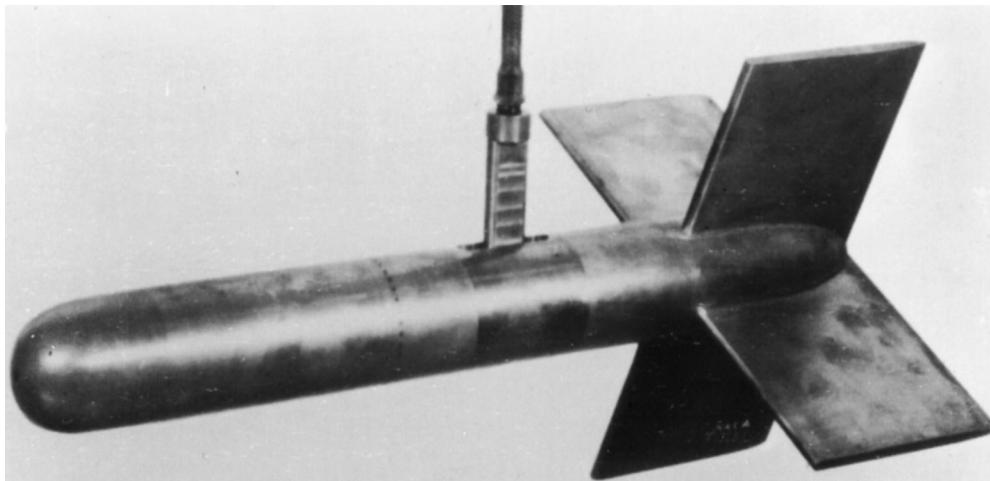


FIGURE 11 Trailing bomb. [From Gracey, W. (1981). "Measurement of Aircraft Speed and Altitude," Wiley-Interscience, New York.]

below the aircraft where the local pressure approximates the free-stream static pressure. Although the bomb senses the pressure below the aircraft, the pressure measured by the bomb system is the pressure at the flight altitude because the pressure is measured by an instrument in the aircraft.

VI. AIR DATA MEASUREMENTS IN THE HYPERSONIC FLIGHT ENVIRONMENT

The current generation of unmanned, trans-atmospheric, reusable launch vehicles—the X33 and X34—and the space station crew return vehicle—the X-38—all feature flight control and guidance systems which require measurement of the aircraft's air data state during the re-entry, approach, and landing portions of the flight trajectory. During these flight phases, Mach number, airspeed, and the angles-of-attack and sideslip are used to ensure that the vehicle has adequate energy to reach the runway and to enhance the flying qualities by compensating for wind conditions. Also, for the X-33, minimizing the angle of attack and angle of sideslip during the launch reduces the loads on the airframe.

A. Measurement Requirements

Inertial guidance systems are not acceptable for the air data measurement task because they do not account for wind conditions. Similarly, air data measurements that are derived using conventional intrusive probes described in the previous sections are unacceptable for these applications because of the severe heating environments encountered during the vehicle re-entry. The space shuttle overcame these difficulties using a complex design that featured de-

ployable probes that were "rolled-out" of the vehicle after the high-heat re-entry conditions were completed. These probes were undesirable for several reasons, including difficulty integrating them into the vehicle structure, complexity of maintaining the probe mechanisms, increased weight, and lack of an accurate method for measuring the angle of sideslip.

B. Flush Airdata Sensing System

To circumvent these problems, the flush air data sensing (FADS) system concept was developed. The FADS concept, where air data are inferred from nonintrusive surface-pressure measurements, does not require probing of the local flow field to compute air data parameters. This innovation allows the extreme hypersonic heating caused by the small radius of a flow-sensing probe to be avoided and extends the useful range of the air data measurement system to the hypersonic flow regime.

Because the FADS system does not "probe" the flow field, but instead uses the natural contours of the forebody, the flow field is much cleaner and easier to calibrate. An additional advantage of the FADS system is that it offers the potential to sense air data on ascent, an option not available to the probe-based system. The system has no movable parts, has minimal maintenance requirements, and is directly integrated onto the nose-cap heat shield.

C. Background

The FADS design builds on work that originated in the early 1960s with the X-15 program (Cary and Keener), continued at NASA Langley (Siemers et al.,; [While, 1983](#)) and Dryden Flight Research Centers ([Larson et al., 1987, 1990](#)) in the 1970s and 1980s, and recently concluded

with flight testing of an onboard real-time system in the early 1990s (Whitmore et al., 1990, 1996). For early real-time FADS algorithms (Whitmore et al., 1990, 1996), surface-pressure measurements were related to the desired air data parameters using a calibrated aerodynamic model derived from the modified Newtonian flow theory (Anderson, 1989). The model captures the salient features of the local flow but is simple enough to be invertable in real time.

D. Solving the Flow Field: The FADS “Triples” Algorithm

An algorithm was developed for the X-33, X-34, and X-38 space vehicles by taking strategic combinations of three sensor readings to analytically decouple the angles-of-attack and sideslip from Mach number, dynamic pressure, and static pressure. This innovation allows for the development of an estimation algorithm whose solution speed is superior to conventional nonlinear regression algorithms and whose stability characteristics can be analytically predetermined for a given port arrangement.

Detailed disclosures of the FADS estimating algorithms and the associated software are presented in Whitmore et al. (1994, 1998).

E. Redundancy Management

Because the air data measurements are critical, redundancy is an essential part of the system. Each flight-critical measurement subsystem must have a fail-operational (Fail-Op) capability. That is, the subsystem can tolerate one failure anywhere in the system software or hardware and still produce a usable result. Typically, this Fail-Op capability is achieved by installing tri-redundant systems, each operating along independent paths. The results of the three systems are then “voted” and the median value signal is selected as the most reliable.

The FADS system has an inherent built-in redundancy. At least five independent pressure measurements must be available to derive the entire air data state vector. Adding additional sensors mitigates the noise sensitivity and adds redundancy and accuracy to the system. Typically, at least seven sensors are used in the FADS system. The FADS design exploits the built-in redundancy of the pressure port matrix to achieve Fail-Op capability with dual-redundant system hardware. This innovation allows for considerable savings in up-front costs.

This redundancy management scheme uses two independent hardware paths and “selects” the system with the best overall fit consistency. That is, the system with the smallest mean-squared fit error (MSE) between the FADS model predictions and the actual pressure measurements

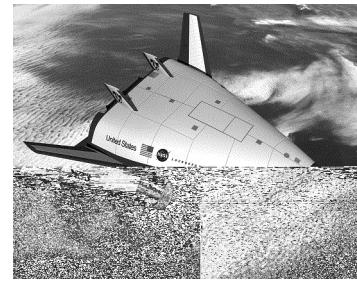


FIGURE 12 Artist's concept of the X-33 vehicle. [From the Dryden Flight Research Center photo archive, EC96-43631-2. Computer art by John Fassanito and Associates.]

is selected as the “safe” system. In general, this MSE fit-error is a good indicator of the overall system health.

F. Summary

The FADS system architecture and algorithms presented here are used as a flight-critical part of the real-time avionics systems for the X-33 (see Figs. 12 and 13) and X-34 advanced launch-systems demonstrators, and the X-38 space station crew recovery vehicle. The FADS design utilizes a matrix of pressure orifices on the vehicle nose to estimate air data parameters. Because the FADS system is nonintrusive, it is readily applicable to hypersonic flow environments. The built-in redundancy of the FADS system allows a Fail-Op capability that is achieved with dual-redundant instead of tri-redundant measurement hardware. This innovation results in a considerable overall cost savings.

VII. ALTIMETRY

Altimeters are used for the measurement of height during takeoff and landing operations, for the clearance of terrain obstacles during en route operations, and for the vertical separation of aircraft.

A. Altimeters

The mechanisms of the altimeters used in service operations are either mechanical or electrical (servoed systems or pressure-transducer systems). With the electrical systems, corrections for the position errors of the static-pressure installation are applied automatically so that the displayed altitude is the correct value.

An example of a mechanical-type altimeter is shown in Fig. 14. This instrument has two scales: (1) an altitude scale with the altitude displayed by a rotating pointer and a revolving drum, and (2) a barometric scale with atmospheric pressure displayed on an adjustable drum. The mechanisms that rotate the pointer drum display and the

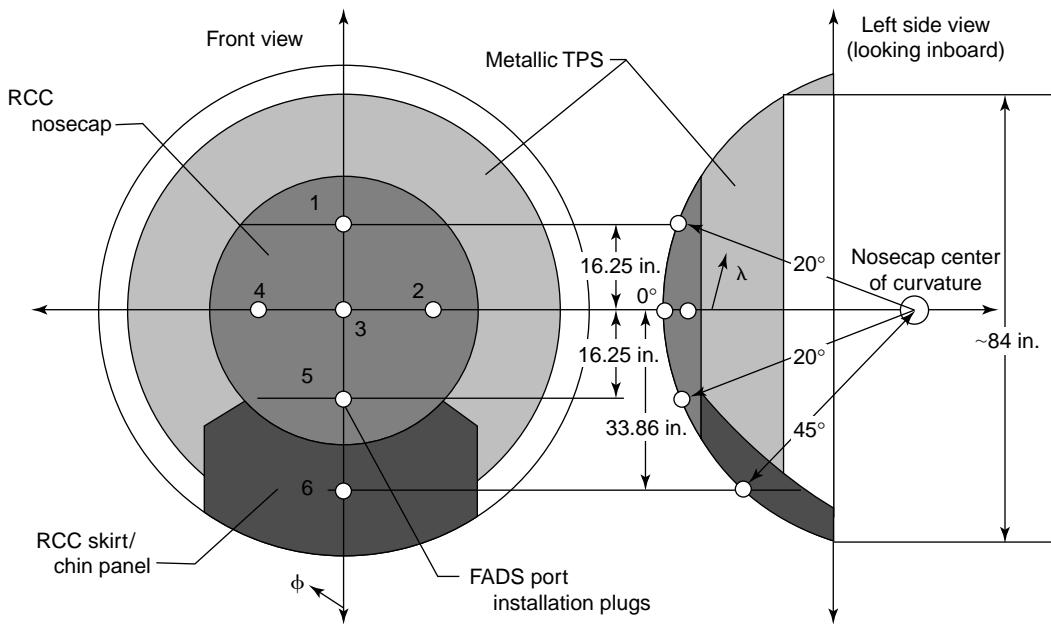


FIGURE 13 X-33 surface port locations. [From Whitmore *et al.* (1998).]

barometric sub-dial are interconnected so that an adjustment of the sub-dial for a given pressure increment produces a rotation of the altitude display to a corresponding increment in altitude. With the altitude display, the pointer rotates one revolution per 1000 ft, while the two-digit drum rotates to indicate 1000-ft and 10,000-ft increments.

B. Barometric-Scale Settings

When the barometric sub-dial is set at 29.92 in. Hg (the sea-level pressure in the standard atmosphere), the altimeter indicates pressure altitude above sea level. This pressure altitude differs from the geometric height whenever

the sea-level pressure or the temperature gradient of the atmosphere differs from the standard values. To account for variations in atmospheric pressure, the barometric sub-dial can be adjusted so that the altimeter indicates either the elevation of the airport or zero height at the airport elevation. Thus, in service operations, the barometric sub-dial may be set at one of three settings that are assigned the following "Q" signals:

- QNE Barometric subdial set at 29.92 in. Hg
- QNH Barometric subdial setting for altimeter to indicate elevation of airport
- QFE Barometric subdial setting for altimeter to indicate zero at the airport

The QNH settings are used by all aircraft for takeoff and landing and for the vertical separation of aircraft at altitudes below 18,000 ft. The QFE settings are used by some airline operators during landing approaches to provide a cross-check with another altimeter set to QNH. The QNE setting is used by all aircraft for vertical separation at altitudes above 18,000 ft.

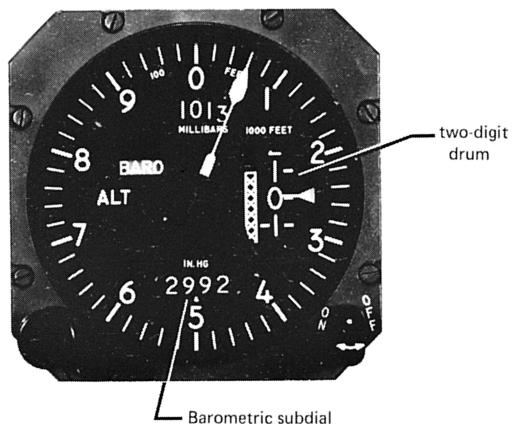


FIGURE 14 Drum pointer altimeter. [Courtesy of Kollsman Instrument Co.]

C. Effects of Pressure and Temperature

The effect of variations in pressure and temperature gradient when the barometric sub-dial is set to QNH can be illustrated with the two hypothetical atmospheric conditions shown in Fig. 15. The curve to the right in both graphs represents the pressure–height relation in the standard

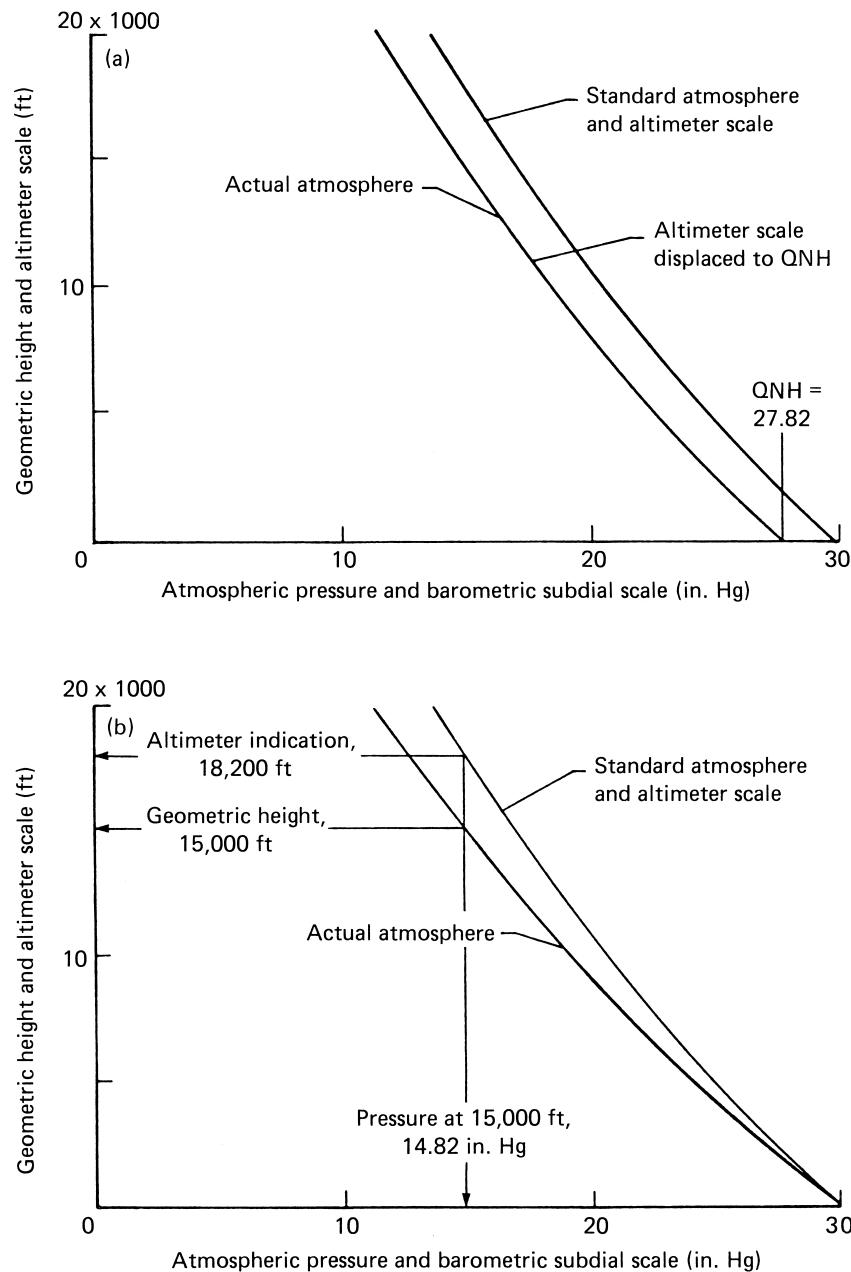


FIGURE 15 Two hypothetical pressure–height variations in the atmosphere. (a) Sea-level pressure below standard and temperature gradient standard; (b) sea-level pressure standard and temperature gradient below standard. [From Gracey, W. (1981). "Measurement of Aircraft Speed and Altitude," Wiley–Interscience, New York.]

atmosphere. Since the barometric scale and the altitude scale of the altimeter have the same relation, an identical curve, representing the two altimeter scales, can be thought to lie on top of the atmospheric curve. Thus, the abscissa of the charts can be labeled barometric sub-dial as well as atmospheric pressure, and the ordinate can be labeled altimeter scale as well as geometric height.

The curve to the left in Fig. 15a represents an atmospheric condition in which the temperature gradient is standard and the sea-level pressure is 27.82 in. Hg. For this condition, the altimeter indicates 2000 ft if the barometric scale is set at 29.92 in. Hg. When the scale is adjusted to 27.82 in. Hg, the altimeter scale curve is moved down until it intersects 27.82 in. Hg on the zero-height axis. The

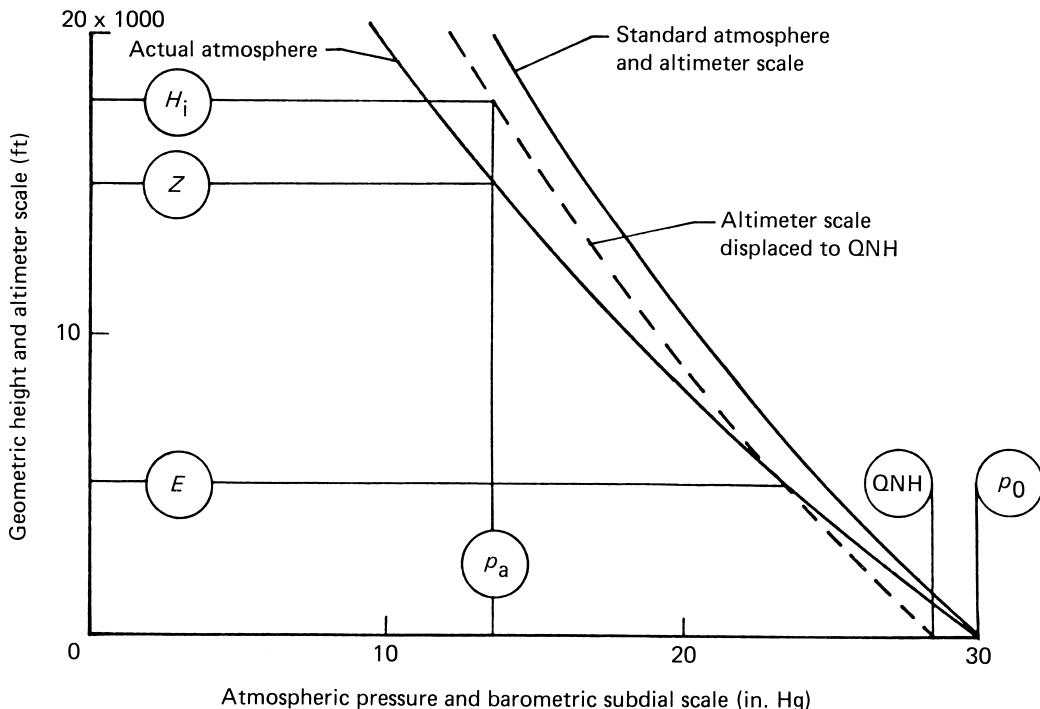


FIGURE 16 Pressure–height variation in an atmosphere in which the sea-level pressure is standard and the temperature gradient is below standard. Altimeter at airport-elevation E is set to the QNH value at the elevation. p_0 is pressure at sea level; Z is the geometric height of the airplane; P_a is the pressure at height Z ; and H_i is the height indicated by the altimeter at height Z . [From Gracey, W. (1981). “Measurement of Aircraft Speed and Altitude,” Wiley–Interscience, New York.]

altimeter pointer will then indicate zero, and the altimeter will indicate geometric height throughout the altitude range.

The curve to the left in Fig. 15b depicts an atmospheric condition in which the sea-level pressure is standard and the temperature gradient is below standard. For this condition, the altimeter indicates zero height at sea level when the barometric scale is set at 29.92 in. Hg (the existing sea-level pressure). At heights above sea level, however, the altimeter indications are higher than the geometric heights. For example, if the altimeter is taken to a height of 15,000 ft, where the existing pressure is 14.82 in. Hg, the altimeter will indicate 18,200 ft (as shown by the intersection of this pressure with the altimeter scale curve).

When the airport elevation is at sea level, the QNH value is the same as the existing sea-level pressure. When the airport elevation is an appreciable height above sea level, however, the QNH value differs from the sea-level pressure whenever the temperature gradient differs from that in the standard atmosphere. This difference can be illustrated by the example shown in Fig. 16. For the case shown, the airport elevation is 5000 ft, the sea-level pressure is 29.92 in. Hg, and the temperature gradient is below

standard. When an altimeter at the airport is adjusted to indicate 5000 ft, the barometric scale indicates 28.30 in. Hg (as shown by the intersection of the altimeter scale curve with the zero-height axis). For this case, therefore, the barometric sub-dial indicates a QNH value that is different from the actual pressure at sea level.

D. Operational Aspects

The QNH setting is used for terrain clearance and for the vertical separation of aircraft up to 18,000 ft. In practice, the pilot adjusts the barometric scale before takeoff until the altimeter indicates the elevation of the airport (QNH value). Before landing at his destination, the pilot resets the barometric scale to the existing QNH value for that area so that the altimeter indicates the elevation of that airport when the aircraft lands. The current QNH settings are measured at the airport weather stations and are reported to the pilots by radio.

During cross-country flights where height information is needed for terrain clearance, particularly in mountainous regions, pilots are required to reset the barometric scales continuously to the QNH values reported by stations along the route. In mountainous areas, the minimum

en route altitude specified by civil regulations is 2000 ft above the highest peak in the region. When the temperature is lower than standard, the geometric height will be lower than the altimeter indication. Studies have shown, however, that even for the lowest recorded temperatures, the difference between the altimeter indication and the geometric height is less than half of the required clearance above the mountains.

For the vertical separation of aircraft, the separation minima are 1000 ft for altitudes up to 29,000 ft and 2000 ft for altitudes above 29,000 ft, respectively. At altitudes above 18,000 ft, the altimeters of all aircraft are set to the same setting (QNE). With this setting, the altimeters in the aircraft above any given point on the earth are referenced to the same pressure. If the reference pressure changes, the flight level of each of the aircraft moves up or down by the same amount, so that the relative separation remains the same (assuming the temperature gradient of the air is standard). If the temperature gradient varies from the standard, the distance between the flight levels decreases when the gradient is below standard and increases when the gradient is above standard. The decrease in separation due to below-standard gradients, however, is generally small compared with the vertical separation minima.

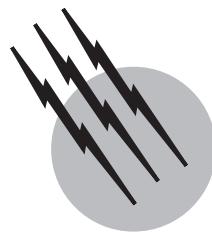
To provide a cross-check of the approximate indications of geometric height provided by altimeters set to the QNH value, some aircraft are equipped with a radio-radar altimeter that provides a direct measure of geometric height above the ground. Since the radio altimeter is more accurate near the ground and the radar altimeter more accurate at high altitudes, the radio altimeter is used for takeoff and landing operations and the radar altimeter is used for terrain clearance at high altitudes.

SEE ALSO THE FOLLOWING ARTICLES

AIRCRAFT AVIONICS • AIRCRAFT INSTRUMENTS • AIRCRAFT PERFORMANCE AND DESIGN • AIRPLANES, LIGHT

BIBLIOGRAPHY

- Anderson, J. D., Jr. (1989). "Hypersonic and High Temperature Gas Dynamics," McGraw-Hill, New York.
- Cary, J. P., and Keener, E. P. "Flight Evaluation of the X-15 Ball-Nose Flow Direction Sensor as an Air Data System," NASA TN D-2923.
- Gracey, W. (1981). "Measurement of Aircraft Speed and Altitude," Wiley- Interscience, New York.
- Larson, T. J., Whitmore, S. A., Ehernberger, L., Johnson, J., Blair, J., and Siemers, P. M., III (1987). "Qualitative Evaluation of a Flush Air Data System at Transonic Speeds and High Angles of Attack," NASA TP-2716.
- Larson, T. J., Moes, T. R., and Siemers, P. M., III (1990). "Wind Tunnel Investigation of a Flush Air Data System at Mach Numbers From 0.7 to 1.4," NASA TM-101697.
- Siemers, P. M., III, Wolf, H., and Henry, M. W. "Shuttle Entry Air Data System (SEADS)—Flight Verification of an Advanced Air Data System Concept," AIAA Paper 88-2104.
- While, D. M. (1983). "Shuttle Entry Air DataSystem (SEADS) Hardware Development," Vol. 1, "Summary," NASA CR 166044, January.
- Whitmore, S. A., Moes, T. R., and Larson, T. J. (1990). "Preliminary Results from a Subsonic High Angle-of-Attack Flush Air Data Sensing (Hi-FADS) System: Design, Calibration, and Flight Test Evaluation," NASA TM-101713.
- Whitmore, S. A., Davis, R. A., and Fife, M. J. (1994). "Flight Demonstration of a Real Time Flush Air Data Sensing (RT-FADS) System," AIAA Paper 94-3433.
- Whitmore, S. A., Davis, R. J., and Fife, J. M. (1996). "In-flight demonstration of a real-time flush air data sensing system," *AIAA J. Aircraft* **33** (5), 970–977.
- Whitmore, S. A., Cobleigh, B. R., and Haering, E. A. (1998). "Design and Calibration of the X-33 Flush Airdata Sensing System, AIAA Paper 98-0201.



Computational Aerodynamics

David A. Caughey

Cornell University

- I. Introduction
- II. Fluid Mechanical Background
- III. Algorithmic Aspects
- IV. Concluding Remarks

GLOSSARY

Artificial viscosity Term added to a numerical approximation to provide an artificial (i.e., nonphysical) dissipative mechanism to prevent round-off errors from accumulating and destroying the accuracy of the solution.

CFL condition Constraint, first described by Courant, Friedrich, and Lewy, that limits the size of the time step by which an explicit method may advance the numerical solution of an initial-value problem to be consistent with the physics of local wave propagation.

Conservation form Divergence form of a system of partial differential equations describing conservation laws, important when solutions containing discontinuities (shocks) are to be computed.

Direct numerical simulation (DNS) Solution procedure for the Navier-Stokes equations in which all scales of turbulent eddies are fully resolved, resulting in an exact solution for turbulent flow.

Euler equations Equations describing the inviscid flow of a compressible fluid. These equations constitute a hyperbolic system of partial differential equations; the Euler equations are nondissipative, and weak solutions containing surfaces of discontinuity (which approxi-

mate the behavior of shock waves) must be allowed for many practical problems.

Explicit method Numerical method for solving initial-values problems in which the solution at a given point and time depends explicitly only upon the solution at preceding time levels.

Implicit method Numerical method for solving initial-values problems in which the solution at a given point and time depends upon the solution at neighboring points at the same time level, as well as upon the solution at preceding time levels.

Inviscid flow Approximation to the fluid-flow equations, useful at large Reynolds numbers, in which the viscous, or internal friction, forces acting within the fluid are neglected.

Large-eddy simulation (LES) Solution procedure for the Navier-Stokes equations in which the largest, energy-containing scales of a turbulent flow are resolved, while the smaller scales are modeled.

Mach number Ratio of flow velocity to speed of sound; a measure of the importance of compressibility in determining the behavior of fluid flows.

Mesh generation The process of generating a volume-filling mesh of hexahedral or tetrahedral cells as a basis for the description of the solution to a nonlinear field

problem. Often a pacing item for the application of computational techniques to fluid-flow problems.

Reynolds-averaged Navier-Stokes (RANS) equations

Equations for the mean properties in a turbulent flow obtained by decomposing the fields into mean and fluctuating components and averaging the Navier-Stokes equations. Solutions of these equations for the mean properties of the flow require knowledge of various correlations (the Reynolds stresses) of the fluctuating components.

Shock capturing Numerical method in which shock waves (and other discontinuities) are incorporated into the solution by smearing them out with artificial dissipative terms so that the solution remains continuous, although with very steep gradients near the shocks.

Shock fitting Numerical method in which shock waves (and other discontinuities) are treated as discontinuities and are explicitly fitted into the solution as internal boundaries, across which specified jump relations are satisfied.

Transonic flow Fluid flow that contains regions of both subsonic and supersonic flow velocities; usually occurs at flight Mach numbers near unity.

Turbulence model Phenomenological model used in Reynolds-averaged Navier-Stokes equations to relate the turbulent stresses to the mean flow properties.

Turbulent flow Flow in which unsteady fluctuations play a major role in determining the effective mean stresses in the field; regions in which turbulent fluctuations are important inevitably appear in fluid flow at large Reynolds number.

Upwind method Numerical method for CFD in which spatial asymmetry is introduced into the difference stencil to introduce dissipation into the approximation, thus stabilizing the scheme. This is a popular mechanism for the Euler equations, which have no natural dissipation, but also is effective for the Navier-Stokes equations, especially at high Reynolds number.

COMPUTATIONAL AERODYNAMICS is concerned with the development and application of numerical techniques to compute solutions to practical aerodynamic problems, such as the design of aircraft, and to pursue fundamental research into the behavior of fluid flow, especially in the turbulent regime. The general subject area includes the development of efficient numerical algorithms and their implementation and use on high-speed, digital computers. Applications include the prediction of aerodynamic forces and heat transfer rates to vehicles in flight. These computational methods have revolutionized the aerodynamic design of flight vehicles.

I. INTRODUCTION

A. Scope of Computational Aerodynamics

In 1946 John Von Neumann, a remarkable pioneer in a variety of fields spawned by the development of high-speed, digital computers, wrote the words:

“Indeed, to a great extent, experimentation in fluid dynamics is carried out under conditions where the underlying physical principles are not in doubt, where the quantities to be observed are completely determined by known equations. The purpose of the experiment is not to verify a proposed theory but to replace a computation from an unquestioned theory by direct measurements. Thus, wind tunnels are, for example, used at present, at least in part, as computing devices of the so-called analogy type . . . to integrate the nonlinear partial differential equations of fluid dynamics.”

Since the time these words were written, the widespread use of computers in science and engineering has had a major impact upon the design practices of modern engineers as well as upon the range of problems that is open to attack by scientists. As suggested by the above quotation, the engineering science of aerodynamics has been a forerunner in the development of computational methods. In this article, the major subject areas of computational aerodynamics will be presented, with a focus on the algorithmic approaches that have resulted in important practical advances in the state of the art. In order to describe these methods, and the algorithmic ideas upon which they are based, it is necessary to introduce enough of the relevant physics of fluid mechanics that the magnitude of the task, and the scope of the achievements to date, can be appreciated.

B. Relation to Wind Tunnel Testing

The time-honored method of aerodynamic evaluation is, of course, the wind tunnel. When used for aerodynamic design purposes (as opposed to its role in fundamental fluid mechanical research), the wind tunnel provides measurements of the aerodynamic forces and heat transfer rates acting on a model of the vehicle of interest. It is this design role of the wind tunnel, not its role in fundamental research, that Von Neumann had in mind when he referred to wind tunnels as analog computers.

Both wind tunnels and computational techniques have their advantages and limitations. The wind tunnel has the advantage that with proper scaling it is, in principle, a perfect simulation of the complete physics (although this proper scaling is rarely achieved in practice). Computational analyses, on the other hand, can be performed quite quickly and are much less energy intensive than wind tunnel testing.

The use of forces measured on a model mounted in a wind tunnel to predict forces on a full-scale vehicle is based upon modeling principles which guarantee that the forces will scale in a particular way if certain dimensionless parameters are the same for both the model and full-scale flows. For aircraft problems the relevant dimensionless parameters are the Reynolds number, which characterizes the relative importance of viscous effects, and the Mach number, which characterizes the relative importance of the compressibility of the fluid (usually air). (Both of these parameters will be defined in later sections.) It generally is not possible to match one or both of these parameters with the corresponding flight parameters in a given test, so great care and considerable insight are required to extrapolate the effects of the mismatched parameter (or parameters). In addition, the wind tunnel usually is used to approximate the flow in an unbounded medium using a flow of limited extent, and the degree to which the walls of the wind tunnel affect the flow in the vicinity of the model is another potential source of error that must be carefully assessed. Of course, these wind tunnel wall effects can be reduced by making the model very small relative to the size of the tunnel, but this makes it harder to match the Reynolds number, which, as will be seen, is proportional to the linear dimension of the model.

Computational methods, on the other hand, are limited by their inability to incorporate enough of the relevant physics into the mathematical model, while still being able to perform the computation with a reasonable expenditure of computer resources, and uncertainties about the degree to which numerical error affects the accuracy with which the equations actually are solved. The constraint on the physical models introduces an uncertainty that again calls for sound judgment to determine whether the missing physics might play an important role. In addition, it often is necessary to approximate the usually complex geometries even in a numerical calculation, and the degree to which this geometrical approximation may compromise the accuracy of the results is another source of uncertainty. The goal of computational aerodynamics is to bring increasingly accurate physical models of the fluid mechanics to bear on problems of ever greater geometrical complexity during engineering design studies, and much progress has been made toward this goal in the past three decades.

The present-day roles of wind tunnel testing and computational aerodynamics are complementary. As increasingly accurate and economical computational techniques are developed, they are used in studies to identify promising candidate designs. These are then tested in the wind tunnel to make certain that any approximations made in the computation are warranted. The best designs are frequently then used as spring boards for new parametric

studies, which are again performed on the computer and validated in the wind tunnel. The net result is that, with approximately the same amount of wind tunnel testing, a much better final design is achieved when computational methods are used to screen candidate configurations. As computational methods continue to evolve, the testing of preliminary designs increasingly will be replaced by computation, but it is likely that final designs will continue to be tested in the wind tunnel for the foreseeable future.

II. FLUID MECHANICAL BACKGROUND

In this section, the fluid mechanical background necessary to understand and appreciate the algorithmic aspects of computational aerodynamics will be described.

A. The Navier-Stokes Equations

As may be inferred from the quotation of John Von Neumann, which was introduced at the beginning of this article, aerodynamics is virtually unique within the engineering sciences in having a generally accepted mathematical framework for describing most problems of practical interest. Such diverse problems as the high-speed flow of air past an airplane wing, the motions of the atmosphere responsible for our weather, and the unsteady air currents associated with the flapping of the wings of a housefly all are described by a set of partial differential equations known as the Navier-Stokes equations. These consist of a system of partial differential equations that express the fundamental conservation laws upon which the science of fluid mechanics is based. These include the conservation of mass, stating that fluid is neither created nor destroyed in any arbitrary region of space; momentum equations that express Newton's second law for the motion of the fluid; and an energy equation based on the first law of thermodynamics. In both the momentum and energy equations, the further approximation that the state of stress within the fluid is linearly related to the rate of strain also is made, but this is widely accepted as an excellent approximation in most instances. These equations can be written in a system of Cartesian coordinates (x, y) as follows. Let ρ and p denote the fluid density and pressure, respectively; u and v the components of the fluid velocity in the x and y directions; and e the total energy per unit volume. For a calorically perfect gas having a ratio of specific heats $\gamma = c_p/c_v$ (a good approximation for air from normal temperatures up to about 1700°C), the pressure is related to the total energy by the equation of state

$$p = (\gamma - 1) \left(e - \rho \frac{u^2 + v^2}{2} \right). \quad (1)$$

The Navier-Stokes equations can then be written in the form

$$\frac{\partial \mathbf{w}}{\partial t} + \frac{\partial \mathbf{f}}{\partial x} + \frac{\partial \mathbf{g}}{\partial y} = \frac{\partial \mathbf{R}}{\partial x} + \frac{\partial \mathbf{S}}{\partial y}, \quad (2)$$

where

$$\mathbf{w} = \{\rho, \rho u, \rho v, e\}^T \quad (3)$$

is the vector of dependent variables and

$$\mathbf{f} = \{\rho u, \rho u^2 + p, \rho uv, (e + p)u\}^T \quad (4a)$$

and

$$\mathbf{g} = \{\rho v, \rho uv, \rho v^2 + p, (e + p)v\}^T \quad (4b)$$

are the flux vectors. The effects of viscosity are governed by the viscous flux vectors

$$\mathbf{R} = \{0, \tau_{xx}, \tau_{xy}, u\tau_{xx} + v\tau_{xy}\}^T \quad (5a)$$

and

$$\mathbf{S} = \{0, \tau_{xy}, \tau_{yy}, u\tau_{xy} + v\tau_{yy}\}^T. \quad (5b)$$

The viscous stresses appearing here are related to the derivatives of the components of the velocity vector by

$$\tau_{xx} = 2\mu \frac{\partial u}{\partial x} - \frac{2}{3}\mu \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) \quad (6a)$$

$$\tau_{xy} = \mu \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) \quad (6b)$$

$$\tau_{yy} = 2\mu \frac{\partial v}{\partial y} - \frac{2}{3}\mu \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right), \quad (6c)$$

where μ is the coefficient of viscosity. In the above, terms representing heat conduction have been omitted from the energy equation for simplicity. Also, here and elsewhere in this article, equations will be written in two space dimensions for the sake of economy; the extension to problems in three space dimensions is straightforward unless otherwise noted.

If the Navier-Stokes equations are nondimensionalized by normalizing lengths with respect to a length L representative of the size of the vehicle past which the flow is to be determined and by normalizing flow properties (such as density, viscosity, and flow velocities) by their values in the undisturbed free stream far from the vehicle, an important *nondimensional parameter*, the Reynolds number

$$\text{Re} = \frac{\rho_\infty V_\infty L}{\mu_\infty}, \quad (7)$$

appears. The viscous stress terms on the right-hand side of Eq. (2) are multiplied by the inverse of Re . Physically, the Reynolds number can be interpreted as a measure of the relative importance of the viscous, or internal friction, forces in determining the behavior of the flow; when the

TABLE I The Hierarchy of Aerodynamic Approximations, after Chapman (1979)

Stage	Model	Equations	Time frame
I	Linear potential	Laplace	1960s
IIa	Nonlinear potential	Full potential	1970s
IIb	Nonlinear inviscid	Euler	1980s
III	Modeled turbulence	Reynolds-averaged Navier-Stokes	1990s
IV	LES	Navier-Stokes (+ subgrid turb. model)	1980s–
V	DNS	Fully resolved Navier-Stokes	1980s–

Reynolds number is large, viscous effects are small except in regions of very large velocity gradients.

The computational resources required to solve the complete Navier-Stokes equations numerically are enormous, especially for problems in which turbulence occurs. A computation that resolves all scales of turbulent motions is called a direct numerical simulation (DNS). In 1971, Howard Emmons of Harvard University estimated the computer time required to solve a simple, turbulent pipe-flow problem, including direct computation of all eddies containing a significant fraction of the turbulent kinetic energy. For a computational domain consisting of approximately 12 diameters of the pipe, the computation of the solution at a Reynolds number of $\text{Re}_D = 10^7$ would require about 10^{17} s on a 1970 main frame computer. Of course, much faster computers are now available, but even for a machine capable of 100 Gigaflops (100×10^9 floating-point operations per second)—representative of the fastest machines currently available—such a calculation would require more than 10,000 years to complete. In fact, it probably is not necessary to resolve all the energy-containing scales in the computation. A great deal of current work is focused on developing suitable dissipative models to remove the required energy of the subgrid scales in these large-eddy simulations (LES) of turbulent flows.

More recently, Dean Chapman of the NASA Ames Research Center has divided problems in computational aerodynamics into four classes, or stages, depending on the degree to which the complete Navier-Stokes equations have been approximated. A computation in which the complete Navier-Stokes equations are solved, including direct simulation of the large-scale turbulent fluctuations, but with the subgrid scale turbulence modeled, is called a Stage IV approximation by Chapman. The computational resources required to solve a Stage IV approximation to the Navier-Stokes equations numerically for the flow past a realistic, but geometrically simplified three-dimensional

wing-body configuration were estimated by Chapman to exceed the capabilities of 1980s generation supercomputers (so-called Class VI supercomputers such as the CRAY 1-S or CYBER 205) by a factor of approximately 4000. Thus, in order to make predictions for practical engineering problems, judicious approximations must be made. The first such simplification that seems reasonable is to model *all* scales of turbulence; this is termed a Stage III computation by Chapman. This can be done, for example, by solving the Reynolds-averaged Navier-Stokes (RANS) equations and modeling phenomenologically the higher order moments required for closure of the system. Such a calculation for a realistic, three-dimensional geometry has been estimated by Chapman to require about 40 times the power of Class VI supercomputers. Such computations became feasible for research purposes around 1990 and are now feasible for limited cases in the design cycle.

B. Boundary Layer Theory

Fortunately, in many aerodynamic problems the effects of viscosity are confined to very thin regions (called boundary layers) in the immediate vicinity of the body surface. Thin boundary layers are a naturally occurring feature of solutions of the Navier-Stokes equations at large Reynolds numbers when the no-slip condition is imposed at solid surfaces (resulting from thermodynamic equilibrium between the surface and the fluid) and the body is sufficiently thin that the viscous layers do not separate from the surface. Outside these viscous layers, the fluid behaves very nearly as a perfect, or inviscid, fluid, which is described by much simpler equations (that will be presented in Section II.C). In addition, since the boundary layers appearing in these flows are thin, this inviscid approximation gives a good approximation to the pressure forces acting on the body surface, and the net forces acting upon the body can be predicted quite accurately. The advantage of this inviscid model is twofold: first, the thin boundary layers no longer need to be resolved in the calculation, and the length scales that need to be resolved in the solution are those determined by the geometry of the body itself; second, since the equations themselves are considerably simpler than the Navier-Stokes equations, they also are significantly less costly (in terms of computer resources) to solve. In addition, numerical methods developed to solve the Navier-Stokes equations must also be capable of solving them accurately in regions where the viscous stresses are not important, i.e., in regions where they reduce to their inviscid counterparts. Thus, algorithms for solving the Navier-Stokes equations often are first tested on the inviscid equations of motion, and it is with attempts to solve these equations describing inviscid flow past aircraft geometries that much of computational aerodynam-

ics is concerned. Therefore, the remainder of this article will concentrate on these aspects, returning to the question of viscous effects and of turbulence only in the closing sections.

C. Inviscid Flow Models

For many design purposes, it is sufficient to represent the flow past an aerospace vehicle as that of an ideal (or inviscid), compressible fluid. This is appropriate for flows at large Reynolds numbers that contain only negligibly small regions of separated flow. The equations describing inviscid flows can be obtained as a simplification of the Navier-Stokes equations in which the viscous terms are neglected altogether. This results in the Euler equations of inviscid flow

$$\frac{\partial \mathbf{w}}{\partial t} + \frac{\partial \mathbf{f}}{\partial x} + \frac{\partial \mathbf{g}}{\partial y} = 0. \quad (8)$$

The Euler equations comprise a hyperbolic system of partial differential equations, and their solutions contain features that do not appear in solutions of the Navier-Stokes equations. In particular, while the diffusion (viscous) terms in the Navier-Stokes equations guarantee that the solutions will be smooth, the absence of these diffusion terms from the Euler equations allows them to have solutions that are discontinuous across surfaces in the flow. Such solutions to the Euler equations must be interpreted within the context of generalized (or weak) solutions, and this theory provides the framework for developing the properties of any discontinuities that may appear. In particular, the jumps in dependent variables (such as density, pressure, and velocity) across such discontinuities cannot be arbitrary but must be consistent with the integral form of the original conservation laws upon which the differential equations are based. For the Euler equations, these jump conditions are called the Rankine-Hugoniot relations. The solution of the Euler equations for steady flows past a realistic, three-dimensional configuration is termed a Stage IIb computation by Chapman, and such computations are routinely performed in the design phase for aerospace vehicles.

If the flow can further be approximated as steady and irrotational, it is possible to introduce a velocity potential Φ such that the velocity \mathbf{q} is given by

$$\mathbf{q} = \nabla \Phi \quad (9)$$

and the steady form of the Euler equations reduces to

$$\frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho v)}{\partial y} = 0, \quad (10)$$

where, from the definition of the potential in Eq. (9),

$$u = \frac{\partial \Phi}{\partial x} \quad \text{and} \quad v = \frac{\partial \Phi}{\partial y}$$

are the velocity components, and the density is given by the isentropic relation

$$\rho = \left\{ 1 + \frac{1}{2}(\gamma - 1)\mathbf{M}_\infty^2 [1 - (u^2 + v^2)] \right\}^{1/(\gamma-1)}, \quad (11)$$

where \mathbf{M}_∞ is the Mach number of the free stream (i.e., the ratio of the velocity to the speed of sound in the free stream). Equation (11) can be used to eliminate the density from Eq. (10), which can then be expanded as

$$(a^2 - u^2) \frac{\partial^2 \Phi}{\partial x^2} - 2uv \frac{\partial^2 \Phi}{\partial x \partial y} + (a^2 - v^2) \frac{\partial^2 \Phi}{\partial y^2} = 0, \quad (12)$$

where a is the local speed of sound.

If a new (s, n) Cartesian coordinate system is introduced with the s axis aligned with the velocity vector, then at any point Eq. (12) can be written as

$$(1 - \mathbf{M}^2) \frac{\partial^2 \Phi}{\partial s^2} + \frac{\partial^2 \Phi}{\partial n^2} = 0, \quad (13)$$

where $\mathbf{M} = \sqrt{u^2 + v^2}/a$ is now the Mach number based on the local flow velocity and speed of sound. This equation shows clearly how the Mach number enters the problem for compressible flows.

Equation (13) is a second-order, quasilinear, partial differential equation whose type depends upon the sign of the coefficient $(1 - \mathbf{M}^2)$. The type of a partial differential equation is intimately related to how information about its solution is transmitted from one point to another, and, in this particular case, it is clear that this flow of information is changed dramatically when the local Mach number changes from subsonic ($\mathbf{M} < 1$) to supersonic ($\mathbf{M} > 1$). When the local Mach number is supersonic, Eq. (13) is said to be of hyperbolic type, and information is transmitted along characteristic surfaces having inclination

$$\left(\frac{dn}{ds} \right)_{\text{char}} = \pm \frac{1}{\sqrt{\mathbf{M}^2 - 1}}$$

relative to the s axis. When the local Mach number is subsonic, the equation is of an elliptic type, and the effect of a change at any point is felt at all other points in the field.

Thus, Eq. (10) contains a mathematical description of the physics necessary to predict the important features of transonic flows. It is capable of changing type, depending upon whether the local Mach number is less than or greater than unity, and the conservation form allows surfaces of discontinuity, or shock waves. Solutions at this level of approximation, while somewhat simpler and more economical than those of the Euler equations, are also considered by Chapman to belong to Stage II and are capable

of accurately describing transonic flows containing weak shock waves.

Finally, if the flow can be approximated by small perturbations to a uniform stream, the potential equation can further be simplified to

$$(1 - \mathbf{M}_\infty^2) \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0 \quad \text{for } |u^2 + v^2 - 1| \ll 1, \quad (14)$$

where ϕ is the perturbation velocity potential defined as

$$\Phi = x + \phi \quad (15)$$

and the velocity of the free stream is assumed to be unity. Further, if the flow can be approximated as incompressible (i.e., the Mach number is everywhere negligibly small), then Eq. (10) reduces to

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0 \quad \text{for } \mathbf{M}_\infty \ll 1. \quad (16)$$

Since the speed of sound in the atmosphere (under standard sea level conditions) is approximately 340 m/s (760 mph), the effects of compressibility will be relatively unimportant for flight speeds of less than about 100 m/s (220 mph). This corresponds to a flight Mach number of about 0.3, and since it is the square of the Mach number that appears in Eq. (14), compressibility corrections are on the order of 10% or less.

Since Eqs. (14) and (16) are linear, superposition of elementary solutions can be used to construct more general solutions. Such surface-singularity methods—called panel methods in aerodynamic applications—have been highly developed for the calculation of flows past quite complex, three-dimensional configurations on relatively modest computers. These methods are included in Chapman's Stage I.

The results of any of the methods corresponding to Stages I and II can be improved by adding corrections to account for the small, but finite, thickness of the boundary layer. For flows having limited regions of separation, such patched solutions are capable of accurately predicting surface pressure distributions (including drag forces), even for transonic flows containing shock waves. The behavior of the flow in the boundary layer can be predicted once the external inviscid flow is known. Computational techniques also are required for predicting the flow within the boundary layer, but space constraints do not allow their description in this article.

III. ALGORITHMIC ASPECTS

The earliest calculations for compressible flows were based, necessarily, upon linearizations of the inviscid

models presented in Section II.C. The hodograph method of analysis uses the specialized fact that, in two dimensions, if the velocity coordinates and physical coordinates are interchanged in Eq. (12), the resulting equation is linear. This hodograph transformation results in a formulation in which one finds the coordinates in the physical plane as a function of the velocity components. Although the equation is linear, the formulation of boundary conditions is sufficiently difficult that only problems with very simple geometries (such as the flow in a planar jet or past a wall having a sharp corner) can be solved without the use of numerical methods. In addition, the formulation does not handle discontinuities easily, since it is now the independent variables that must change discontinuously.

A second form of linearization is achieved by approximating the nonlinear terms in Eq. (10) by some specified function. [There is an interesting analogy with turbulence models that frequently take assumed forms for the dependence of time averages of higher order (or nonlinear) moments of the fluctuating quantities.] In the case of compressible flows, the approach must be of severely limited applicability, since the assumed form of the nonlinear terms implies *a priori* knowledge of where the sonic lines and shocks must appear in the solution.

In order to treat problems of sufficient geometrical complexity to be of practical value, numerical methods are required.

A. The Panel Method

The earliest numerical methods used for compressible flow problems were based on making compressibility corrections to solutions of the linear potential problem, defined by Eq. (14). In fact, programs for solving these problems were among the first widely used, large-scale numerical methods for engineering analysis. In addition, solutions of Eq. (16) are also of interest for low-speed, or nearly incompressible, flows. The linearity of these equations allows the construction of arbitrarily complex solutions by the superposition of relatively simple (analytically describable) elementary solutions. The elementary solutions used in these analyses correspond to distributions of singular solutions whose forms are specified over each of a number of nonoverlapping subdomains on the surface of the vehicle. The strengths of the singularities on each of these facets (or panels) is determined by satisfying the boundary condition that there be no fluid flux through the surface of the body.

A major advantage of panel methods, relative to the more general analyses that will be described in Section III.B, is that it is necessary to describe (and to discretize into panels) only the surface of the body; thus, a three-dimensional solution—such as that for the flow

past a complete vehicle—requires the solution of a basically two-dimensional problem. The panel method has been developed extensively for practical problems since its first application to three-dimensional problems more than 30 years ago, but since the techniques used to solve the potential equation in this way are quite different from those used to solve nonlinear problems, and these methods are not readily extendable to the nonlinear case, they will not be discussed further here.

B. Nonlinear Methods

1. General Considerations

As is seen in Section II.C, linear methods can never predict truly transonic phenomena; calculations of transonic flow past aircraft configurations must be based on at least the full potential equation or the Euler equations. For such nonlinear problems, it is necessary to describe the solution throughout the entire domain (not just on the body surface, as in the panel method) and to use some sort of iteration (which may, or may not, actually approximate the physics of an unsteady flow process) to drive the flow variables toward the solution of the nonlinear equations. Solutions for both the full potential and Euler equations generally are based on finite-difference or finite-volume techniques. In both these techniques, a grid, or network, of points is distributed throughout the flow field, and the solution is represented by its values at these points. In a finite-difference method the derivatives appearing in the original differential equation are approximated by finite differences of the values of the solution at neighboring points (and times if the flow is unsteady), and substitution of these into the differential equation yields a system of algebraic equations relating the values of the solution at neighboring grid points. In a finite-volume method, the unknowns are taken to be representative values of the solution in the control volumes formed by the intersecting grid surfaces, and the equations are constructed by balancing fluxes across the surfaces of each control volume with the rates of change of the variables within the control volume. The algebraic equations relating the values of the solution in the neighboring cells are usually very similar in appearance to those resulting from a finite-difference formulation. A third class of numerical methods, which is more highly developed for elliptic problems (and is widely used in structural analysis), is called the finite-element method. In that method, the solution is represented using simple interpolating functions within each mesh cell, or element, and the equations for the nodal values are obtained by integrating a variational or residual formulation of the equations of motion over the elements. Such methods are less widely used for compressible flows,

although significant progress has been made in the past decade.

Since the potential equation and the Euler equations are nonlinear, the algebraic equations resulting from either a finite-difference or finite-volume formulation also are nonlinear, and a scheme of successive approximation usually is required to solve them. These equations are highly local in nature, however, and efficient iterative methods usually are available to solve them.

Treatment of complex geometries. Although the equations of fluid mechanics summarized in Section II have been written in a simple Cartesian coordinate system, their solution for practical engineering problems usually requires that they be expressed on a grid that matches the shape past which the flow is to be computed. Solving the problem on such a *boundary-conforming* (or body-fitted) grid system has the advantages of (1) eliminating the need for cumbersome, and possibly inaccurate or destabilizing, interpolation formulas to enforce the boundary conditions and to obtain the solution on the body surface and (2) allowing the relatively easy and efficient clustering of mesh surfaces near the body surface where gradients in the solution usually are the largest and where high accuracy is required.

The problem may be formulated using either a *structured* or an *unstructured* grid system. Structured grids are those in which the grid points can be ordered in a regular Cartesian structure, i.e., the points can be given indices (i, j) such that the nearest neighbors of the (i, j) point are identified by the indices $(i \pm 1, j \pm 1)$. The grid cells for these meshes are thus quadrilateral in two dimensions and hexahedral (but having nonplanar faces) in three dimensions. Unstructured grids have no regular ordering of the points or the cells, and a connectivity table must be maintained to identify which points and edges belong to which cells. Unstructured grids most often consist of triangular cells in two dimensions and tetrahedral cells in three dimensions, or combinations of these and quadrilateral and hexahedral cells, respectively. In addition, grids having purely quadrilateral (or hexahedral) cells may also be unstructured, e.g., when multilevel grids are used for adaptive refinement.

Implementations on structured grids are generally more efficient than those on unstructured grids, since indirect addressing is required for the latter, and efficient implicit methods often can be constructed that take advantage of the regular ordering of cells in structured grids.

For structured grids, the grid surfaces can be viewed as surfaces of a new, transformed, coordinate system, and the expression of the system of conservation laws in this new coordinate system thus reduces the problem to one of (apparently) Cartesian geometry. Fortunately, it is quite

easy to transform systems of conservation laws to a new coordinate system. The transformation will be described here for the Euler equations [Eqs. (8) of Section II.C]. The local properties at any point of a transformation to a new coordinate system (ξ, η) are contained in the Jacobian matrix \mathbf{J} , which is defined as

$$\mathbf{J} = \begin{pmatrix} x_\xi & x_\eta \\ y_\xi & y_\eta \end{pmatrix} \quad (17)$$

and for which the inverse is

$$\mathbf{J}^{-1} = \begin{pmatrix} \xi_x & \xi_y \\ \eta_x & \eta_y \end{pmatrix} = \frac{1}{h} \begin{pmatrix} y_\eta & -x_\eta \\ -y_\xi & x_\xi \end{pmatrix}, \quad (18)$$

where

$$h = x_\xi y_\eta - x_\eta y_\xi \quad (19)$$

is the determinant of the Jacobian matrix. In these equations, subscripts denote partial differentiation with respect to the corresponding independent variable.

Introduction of the contravariant components of the flux vectors

$$\{\mathbf{F}, \mathbf{G}\}^T = \mathbf{J}^{-1} \{\mathbf{f}, \mathbf{g}\}^T \quad (20)$$

allows the Euler equations to be written in the compact form

$$\frac{\partial h\mathbf{w}}{\partial t} + \frac{\partial h\mathbf{F}}{\partial \xi} + \frac{\partial h\mathbf{G}}{\partial \eta} = 0 \quad (21)$$

when the transformation is independent of time [i.e., the (ξ, η) coordinates are fixed in Cartesian space]. If the coordinate transformation is time dependent (as might be required if the vehicle is deflecting control surfaces or the structure is deforming due to aerodynamic loads), a similar conservation form can also be derived but includes additional fluxes when the surfaces of the control volume are moving through space.

The Navier-Stokes equations can be transformed in a similar manner, although the transformation of the viscous terms is somewhat more complicated and will not be given here. Since the potential equation is simply the continuity equation (the first of the conservation laws comprising the Euler equations), the transformed potential equation is immediately seen to be given by

$$\frac{\partial(\rho h U)}{\partial \xi} + \frac{\partial(\rho h V)}{\partial \eta} = 0, \quad (22)$$

where

$$\{U, V\}^T = \mathbf{J}^{-1} \{u, v\}^T \quad (23)$$

are the contravariant components of the velocity.

The generation of a suitable boundary-conforming coordinate system often is a major obstacle to solving problems of practical interest, particularly when dealing

with very complex geometries. A number of techniques, including conformal mapping, algebraic transformation sequences, and the solution of systems of elliptic or hyperbolic partial differential equations for the mesh coordinates, are used. For very complex geometries, the domain may be subdivided into a number of subdomains, with separate grids constructed in each of the blocks. Such a grid is called a multiblock grid. A boundary-conforming grid used to compute the flow past a two-dimensional airfoil is shown in Fig. 1. This grid was generated using a combined algebraic-conformal mapping procedure. It contains 16,384 cells, but only a fraction of the cells—those nearest the airfoil surface—are shown for the sake of clarity. The complete grid extends to a circle having a radius equal to 30 body lengths, where a uniform freestream boundary condition is imposed on the solution.

As described above, an alternative approach is to abandon the lexicographical ordering of the grid and to use an unstructured grid. Most commonly this is done using cells of triangular (in two dimensions) or tetrahedral (in three dimensions) shape, but other shapes can be used as well. For unstructured grids, the mesh generation step is no longer usefully considered a coordinate transformation, and the equations are developed by computing the fluxes directly in the Cartesian system. There is considerably more overhead (both in storage and in CPU time) for such methods, but the greater ease of mesh generation usually makes this the method of choice for complex geometries, especially in three dimensions. Unstructured grids also can be adapted locally to features of the solution, such as large gradients and shock waves, much more efficiently than structured grids. General methods for the

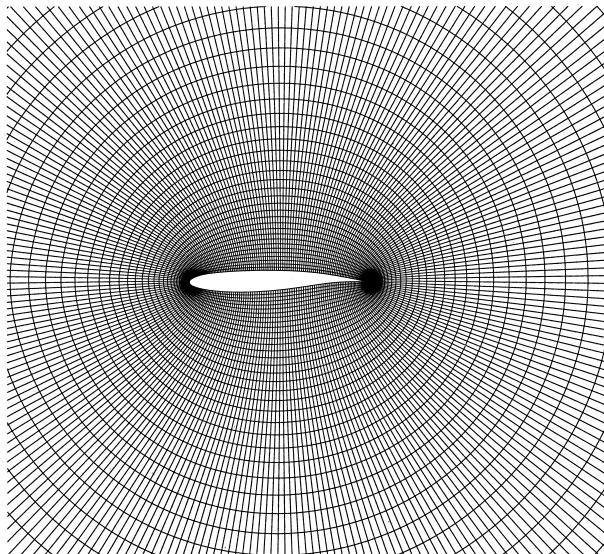


FIGURE 1 Structured grid for an airfoil computation.

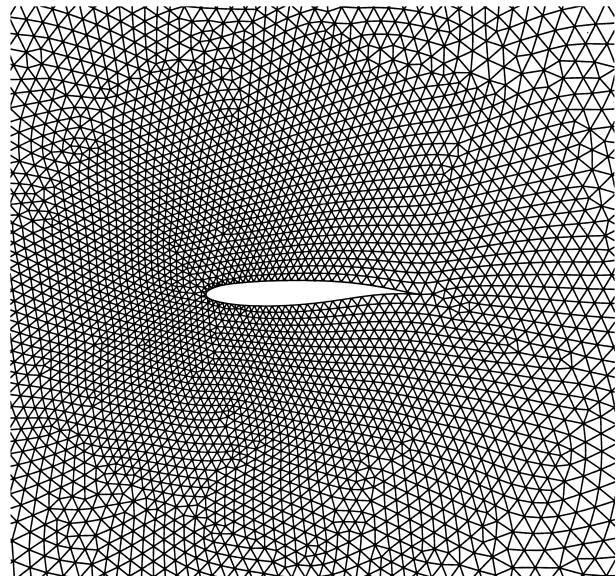


FIGURE 2 Unstructured grid for an airfoil computation.

automatic generation of unstructured grids for very complex, three-dimensional geometries have been developed, based on Delaunay triangulation techniques and so-called *advancing front* methods. An unstructured grid consisting of triangular cells in the vicinity of a two-dimensional airfoil geometry is shown in Fig. 2.

Shock capturing. Most numerical methods for solving compressible flow problems rely on the introduction of artificial (nonphysical) terms in the equations to smear out any shock waves that develop in the solution and allow shocks to be captured naturally by the numerical scheme without any special treatment at points near the shocks. This *shock capturing* approach is in contrast with that of *fitting* the shocks as surfaces of discontinuity, which then must be treated as internal boundaries in the flow calculation, across which the appropriate jump relations must be enforced as an internal boundary (or compatibility) condition. The artificial (or numerical) viscosity added in a shock-capturing scheme acts to smear out the discontinuities that the inviscid theory would predict in much the same manner as molecular viscosity smears out shock waves in the real world. The length scales of these phenomena are, of course, quite different. Under conditions typically found in the atmosphere, shock wave thicknesses are of the order of a few microns (10^{-6} m); under the action of the artificial viscosity of a numerical scheme, the shock thicknesses scale with the grid spacing, which might correspond to a physical distance of 10 cm for a reasonable mesh spacing on a full-scale wing. It is important to realize that even when the Navier-Stokes equations are solved, artificial viscosity usually is necessary when

the solution contains shock waves, since it is impractical to use mesh spacings fine enough to resolve the physical shock structure defined by the molecular viscosity.

2. Transonic Potential Methods

The primary advantage of solving the potential equation, rather than the Euler equations, results from the existence of a velocity potential in the former case, allowing the solution to be described in terms of a single scalar function Φ , rather than the vector of unknowns, and resulting system of equations, in the more general case. The formulation of numerical schemes to solve this single scalar equation is complicated by the fact that, as noted earlier, the equation changes type when the local Mach number changes from subsonic to supersonic and vice versa. (In contrast, the unsteady Euler equations are always of hyperbolic type.) Differencing schemes for the potential equation must, therefore, be *type dependent* (i.e., must change their form depending upon whether the local Mach number is subsonic or supersonic). These methods generally are based on central, or symmetric, differencing formulas, which are appropriate for subsonic flows in which disturbances are free to spread in all directions, with modifications to reflect the directionality of signal propagation in supersonic regions, in which disturbances introduced at a point can be felt only in a limited, conical region downstream of the point. This directional bias can be introduced into the difference equations either by adding an artificial viscosity proportional to the third derivative of the potential Φ or by replacing the density at each point in the supersonic zone by its value slightly upstream of the actual location. Algebraically, the two approaches are equivalent, since this upwinding of the density evaluation also effectively introduces a correction proportional to the third derivative of the potential.

It is important that such artificial viscosity (or compressibility) be added in such a way that the effect vanishes as the mesh spacing is refined. In this way, the numerical approximation will approach the differential equation in the limit of zero mesh spacing, and the method is said to be *consistent* with the original differential formulation. In addition, for flows with shock waves it is important to base the method on the *conservation form* of the potential equation, Eq. (10), rather than the quasilinear form of Eq. (12). This will ensure that the shock jumps are uniquely determined. The shock jump relation corresponding to Eq. (10) is different, however, from the Rankine-Hugoniot condition for the Euler equations. Since the entropy is everywhere conserved in the potential formulation, and since there is a finite entropy jump across a Rankine-Hugoniot discontinuity, it is clear the solutions must be different. For weak shocks, however, the entropy jump given by

the Rankine-Hugoniot condition is of third order in the strength of the shock, so the difference is quite small, and the economies afforded by the potential theory make computations based upon this approximation attractive for many transonic problems.

The nonlinear difference equations resulting from a discrete approximation to the potential equation generally are solved using an iterative, or relaxation, technique. The equations are linearized by computing approximations to all but the highest derivatives from the preceding solution in an iterative sequence, and a correction is computed at each mesh point in such a way that the approximate equations are satisfied identically. It frequently is useful in developing these iterative techniques to think of the iterative process as a discrete approximation to a continuous, time-dependent process. Thus, the iterative process approximates an equation of the form

$$\beta_0 \frac{\partial \Phi}{\partial t} + \beta_1 \frac{\partial^2 \Phi}{\partial \xi \partial t} + \beta_2 \frac{\partial^2 \Phi}{\partial \eta \partial t} = \frac{a^2}{\rho} \left[\frac{\partial(\rho h U)}{\partial \xi} + \frac{\partial(\rho h V)}{\partial \eta} \right]. \quad (24)$$

The parameters β_0 , β_1 , and β_2 can then be chosen to ensure that the time-dependent process converges to a steady state in both subsonic and supersonic regions.

Even with the values of the parameters chosen to provide stable convergence to the steady state solution, many hundreds of iterations may be necessary for the iteration to converge to within an acceptable tolerance of the exact solution to the difference equations, especially when the grid is very fine. This slow convergence on fine grids is a characteristic of all iterative schemes and is a result of the fact that the difference equations are highly local and the elimination of the global low wave number component of the error naturally requires many sweeps through the field. A powerful technique for circumventing this difficulty with the iterative solution of partial differential equations has been developed and applied with great success to aerodynamic problems. The technique is known as the multigrid method and uses the fact that after a few sweeps of any good iterative technique the error remaining in the solution is quite smooth and can be represented accurately on a coarser grid. Application of the same iterative technique on this coarser grid soon makes the error smooth on this grid as well, and the process of coarsening the grid can be repeated again and again until the grid contains only a few cells in each coordinate direction. The corrections that have been computed at all grid levels are then added back into the solution on the fine grid, and the entire multigrid cycle is repeated. The accuracy of the converged solution on the fine grid is determined by the accuracy of the approximation on that grid, since the coarser levels are used only to effect a more rapid convergence of the

iterative process. In theory, for a broad class of problems the work (per mesh cell) to solve the equations in this way is independent of the number of mesh cells, which is the best order estimate one can hope for. In many practical calculations, as few as 10 multigrid cycles may be required.

One of the earliest successes of computational aerodynamics has been the design of so-called *shock-free* configurations for flight in the transonic regime. The transonic drag rise associated with the appearance of shock waves near the wings of aircraft cruising at high subsonic Mach numbers limits the speeds at which subsonic transport aircraft can fly economically. Numerical methods capable of analyzing transonic flows have led to the ability to design airfoils and wings capable of supporting large regions of locally supersonic flow, but maintaining smooth recompression of the flow back to subsonic velocities. Figure 3 shows the surface pressure distribution of one such airfoil (designed by David Korn at the Courant Institute of Mathematical Sciences of New York University) at its design condition of zero angle of attack and 0.75 Mach number. Since the airfoil is cambered (i.e., its upper and lower surfaces are not symmetrical), it produces lift even at zero angle of attack, and since there is an appreciable acceleration of the flow over the upper surface to produce this lift, there is a sizeable pocket of supersonic flow in the vicinity of the upper surface. Figure 3 shows the pressure coefficient, which is a nondimensional measure of the amount by which the pressure differs from its value if the free stream, on the upper and lower surfaces of the airfoil as a function of distance along the airfoil surface. Aerodynamicists conventionally plot negative values of pressure coefficient up, so higher values correspond to lower pres-

sures (and, correspondingly, higher Mach numbers). For comparison, results of computations using both the potential and Euler equations are shown; for this essentially shock-free flow, the two solutions are virtually identical.

Finally, although shock-free flows are achievable, it is important to realize that they usually are quite sensitive to small changes in body geometry, as well as to changes in flight Mach number and/or angle of attack. To illustrate this (and also to show some pressure distributions containing shock waves), the surface pressure distributions are plotted for the same airfoil at Mach numbers slightly below and above the design value in Figs. 4a and 4b. At the lower Mach number, the supersonic pocket breaks into two smaller zones, each terminated by a weak shock wave; at the higher Mach number, the supersonic zone grows significantly larger and is terminated by a single (much stronger) shock. Note the larger differences between the potential and Euler solutions as the shock wave becomes stronger.

While results have been shown here only for two-dimensional computations, transonic potential calculations have become common for three-dimensional wings and wing-fuselage combinations in both industry and government research laboratories, including those of NASA. These methods were heavily used in the design of the Boeing 757 and 767 aircraft and in preliminary design work for the Boeing 777.

3. Euler Equation Methods

As noted earlier, the Euler equations of inviscid, compressible flow constitute a hyperbolic system of partial

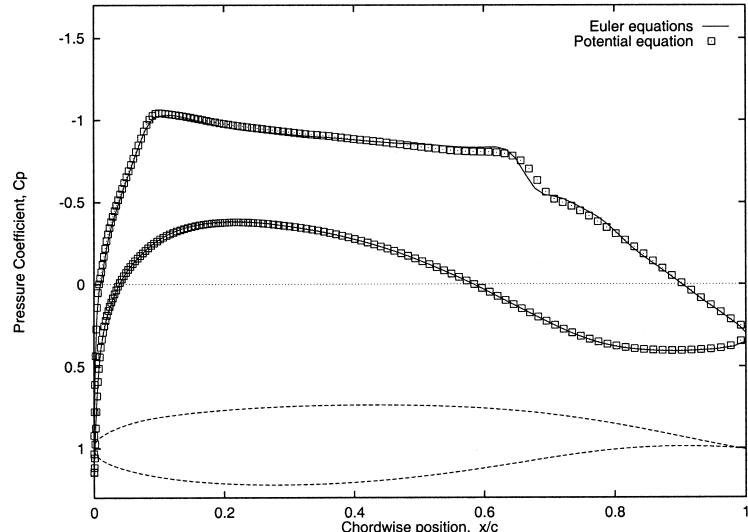


FIGURE 3 Surface pressure distribution on a shock-free airfoil at its design Mach number $M_\infty = 0.75$. Results of both the nonlinear potential and the Euler equations are shown.

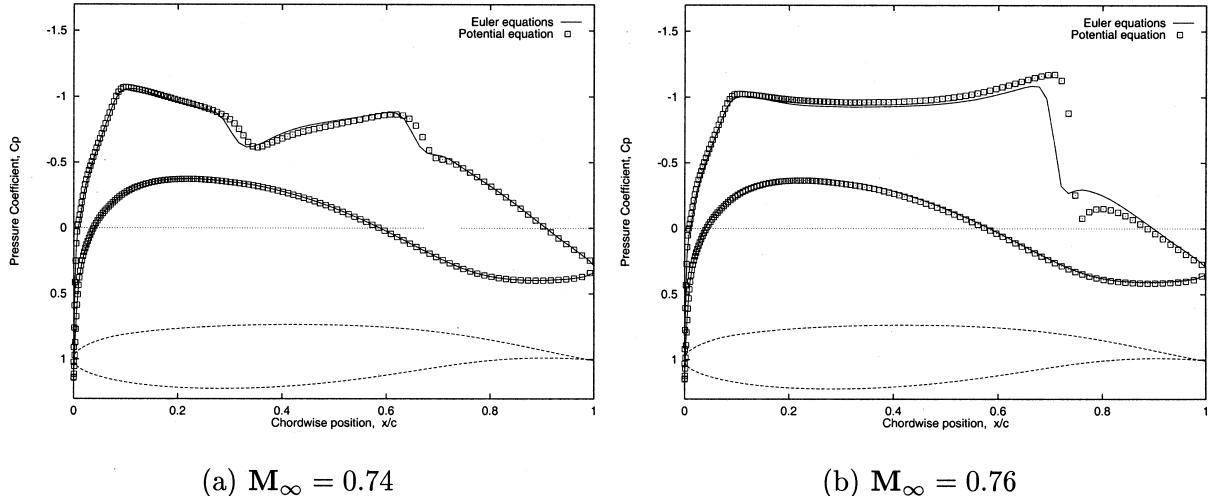


FIGURE 4 Surface pressure distributions for a shock-free airfoil at off-design conditions. Results of both the nonlinear potential and the Euler equations are shown.

differential equations, and numerical methods to solve them for aerodynamic applications rely heavily on the rather well-developed mathematical theory of such systems. As in the case of the nonlinear potential theory, discontinuous solutions (corresponding to shock waves) play an important role. Most methods are also of the shock-capturing type, so it is important to maintain the conservation form of the equations to ensure that the shock strengths are properly determined. In fact, the ideas of shock capturing and the importance of conservation form were developed first within the framework of the mathematical theory of hyperbolic systems of conservation laws. Hyperbolic systems describe the evolution in time of physical systems undergoing some unsteady process. Thus, if the state of the system is known at some initial instant, the hyperbolic system will determine its state at any subsequent time. This feature is frequently used even in calculations for which the solution is steady (i.e., independent of time). In this case, it is the asymptotic solution for large time that is of interest, so the equations are solved for large enough values of the time that the steady state is closely approximated. To maintain the hyperbolic character of the solution, and to keep the numerical method consistent with the physics it is trying to predict, it is necessary to determine the solution sequentially at a number of intermediate time levels between the initial and final (asymptotic) states; such a sequential process is said to be a time marching of the solution.

The simplest hyperbolic partial differential equation is the one-dimensional equation for a single scalar variable

$$\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x} = 0, \quad (25)$$

where $a = \partial f / \partial u$ is real. Although it is important that the numerical implementation be based on this conservation form of the equation, many questions about the behavior of the equation are more easily studied using the quasilinear form

$$\frac{\partial u}{\partial t} + a \frac{\partial u}{\partial x} = 0, \quad (26)$$

where a is assumed constant for purposes of analysis. Most methods for analyzing the behavior of numerical algorithms apply only to equations that are linear, but they provide useful information because stability of the linear equation is a necessary condition for stability of the nonlinear system.

Explicit methods. The simplest practical methods for solving hyperbolic systems are *explicit* in time. That is, the value of the solution is computed at each point at a particular level of time using only the values of the solution at the preceding time level. This is in contrast to *implicit* methods, which may also use neighboring values of the solution at the new time level and thus require the solution of an algebraic system of equations at each time level. This need to solve algebraic systems (which often are similar to the equations resulting from a discretization of an elliptic problem, such as steady, subsonic potential flow) makes implicit methods more costly in terms of computer resources per time step than explicit methods. The size of the time step that can be used in an explicit method, however, is limited by a constraint known as the Courant-Friedrichs-Lowy (or CFL) condition. If Δx is the mesh spacing, then for Eq. (26) the maximum allowable time step Δt for most common explicit schemes is determined by the constraint that the Courant number

$$C = \frac{a\Delta t}{\Delta x} \leq 1. \quad (27)$$

Broadly interpreted, the CFL condition states that the time step must be no greater than the time required for a signal to propagate across a mesh cell. Thus, if the mesh is very fine, the allowable time step also must be very small, and many time steps must be taken to reach the desired asymptotic steady state.

The most widely used explicit methods for hyperbolic systems are of the Lax-Wendroff type. The form of the difference corresponding to one of these methods will be given later in Eqs. (29a) and (29b). At this point, a physical interpretation of their characteristics will be provided by describing the modified equation corresponding to the method when applied to Eq. (26). The modified equation is the differential equation corresponding to the numerical approximation when the leading error terms, proportional to various powers of the mesh spacing, introduced by the approximation are included. For Lax-Wendroff methods applied to Eq. (26), the modified equation is

$$\frac{\partial u}{\partial t} + a \frac{\partial u}{\partial x} = a \frac{\Delta x^2}{6} (C^2 - 1) \left[\frac{\partial^3 u}{\partial x^3} + \frac{3\Delta x}{4} \frac{\partial^4 u}{\partial x^4} \right]. \quad (28)$$

It can easily be shown that derivatives of even order appearing on the right-hand side of Eq. (28) cause changes in the amplitude of the solution, while derivatives of odd order cause changes in the speed at which the solution propagates through space. Since there are no higher order derivatives in the original equation, these terms represent errors. The former are called dissipative (or amplitude) errors, while the latter are called dispersive (or phase) errors. Note that, although there are no dissipative or dispersive terms in the Euler equations, the viscous terms in the Navier-Stokes equations are dissipative in nature. By analogy, the introduction of artificial (or numerical) dissipative terms into a numerical approximation is frequently described as the addition of an *artificial viscosity*.

It is clear from the form of the coefficients in Eq. (28) that both types of errors for the Lax-Wendroff scheme can be made arbitrarily small by reducing the size of the mesh spacing Δx , but that for any finite mesh spacing the errors will remain. In fact, some dissipative error is helpful in preventing unwanted oscillations in the solution (caused by the round-off error introduced by the finite precision arithmetic of any practical computer) from growing larger. In particular, a negative coefficient of the fourth-derivative term will result in exponential damping of these perturbations. This is ensured by requiring that the Courant number C remain less than unity. This is an alternative interpretation of the CFL condition.

The simplest and most efficient form of the Lax-Wendroff scheme, and one that has been widely applied

to aerodynamic problems, is MacCormack's method. If the values at uniformly spaced mesh points i at time level $n\Delta t$ are denoted as u_i^n , MacCormack's method for Eq. (25) can be written as a two-step sequence in which provisional values \bar{u}_i^{n+1} are computed according to

$$\bar{u}_i^{n+1} = u_i^n - \frac{\Delta t}{\Delta x} (f_i^n - f_{i-1}^n) \quad (29a)$$

and then corrected according to

$$u_i^{n+1} = \frac{1}{2} \left[u_i^n + \bar{u}_i^{n+1} - \frac{\Delta t}{\Delta x} (\bar{f}_{i+1}^{n+1} - \bar{f}_i^{n+1}) \right]. \quad (29b)$$

It is easily verified that this is a consistent approximation to the modified Eq. (28). For flows with shock waves, the very small dissipation provided by the term that is third order in the mesh spacing must be augmented by additional artificial viscosity terms in practical calculations.

A popular alternative approach is to treat the spatial and temporal discretizations separately. If the spatial derivative appearing in Eq. (26) is approximated using finite differences, but the time variable is left continuous, there results (for a simple centered spatial difference)

$$\frac{du_i}{dt} + a \frac{u_{i+1} - u_{i-1}}{2\Delta x} = 0. \quad (30)$$

This is a system of ordinary differential equations relating the change in the solution at each point to the values of its neighbors. The well-developed science of numerical methods for solving systems of ordinary differential equations can now be brought to bear on this problem. In particular, conventional multistage Runge-Kutta schemes are well suited to solve the equations arising from this approach for hyperbolic problems (i.e., problems having little or no dissipation). In addition, this approach is attractive for use in conjunction with multigrid methods, as the dissipative characteristics of the time-stepping algorithm can be tailored so that it is an effective smoother of high wave number errors.

Of course, the Euler equations are a system of equations, not a single scalar equation. But, the ideas of this section can easily be extended to systems, as can the implicit methods to be described in Section III.B.3.

Implicit methods. The time-step limitation imposed by the CFL condition often is overly restrictive, especially when only the asymptotic steady-state solution is of interest. In this case, it is desirable to take very large time steps so that the asymptotic steady state is reached with a minimum of computational effort. Time steps larger than those allowed by the condition $CFL \leq 1$ are possible only with the use of implicit methods, but at the cost of having to solve systems of algebraic equations at each time step. For a particular problem, then, the implicit method will

be more economical only if the advantage of the larger time step outweighs the increased cost per time step of the method.

Using the same nomenclature as for MacCormack's scheme, a general implicit method using data at two time levels for a system of equations of the form of Eq. (25) can be written as

$$\mathbf{u}_i^{n+1} = \mathbf{u}_i^n - \Delta t \frac{\theta \delta_x \mathbf{f}_i^{n+1} + (1 - \theta) \delta_x \mathbf{f}_i^n}{2 \Delta x}, \quad (31)$$

where

$$\delta_x \mathbf{f}_i = \mathbf{f}_{i+1} - \mathbf{f}_{i-1}$$

is a central difference operator, and θ is a parameter that varies from 0 for an explicit scheme to 1 for a fully implicit one. For the Euler equations, \mathbf{f} is a nonlinear function of \mathbf{u} , and iteration would normally be necessary to solve Eq. (31) since it is necessary to know \mathbf{u}^{n+1} in order to determine \mathbf{f}^{n+1} . This could be very costly in terms of computer time and usually is avoided by linearizing the flux vector about the previous time level to give

$$\mathbf{f}_i^{n+1} = \mathbf{f}_i^n + \{\mathbf{a}\} \Delta \mathbf{u}_i^n + O(\Delta t^2), \quad (32)$$

where $\{\mathbf{a}\}$ is now a square matrix, called the Jacobian of the flux vector \mathbf{f} with respect to the solution \mathbf{u} , and

$$\Delta \mathbf{u}_i^n = \mathbf{u}_i^{n+1} - \mathbf{u}_i^n$$

is the change in the solution from the n th to the $n + 1$ st time level. This linearization allows the scheme of Eq. (31) to be approximated as

$$[\mathbf{I} + \theta \{\mathbf{a}\}] \Delta \mathbf{u}_i^n + \delta_x \mathbf{f}_i^n = 0, \quad (33)$$

which is a system of linear equations that can be solved for the changes in the solution at each point. Solution of Eq. (33) requires inversion of a system of equations that is block tridiagonal (i.e., the basic structure of the matrix corresponding to the implicit operator is tridiagonal, but the elements are square blocks of dimension equal to the number of equations in the system of differential equations). The blocks are 3×3 , 4×4 , or 5×5 for the one-, two-, and three-dimensional Euler equations, respectively.

A similar procedure can be followed to construct implicit schemes for multidimensional problems. In this case, however, the expense of solving the equations is greatly increased by the fact that the bandwidth of the equations is much larger since the change in the solution in each cell is related to those at its neighbors in each of the coordinate directions. The problem still can be made computationally tractable, however, by further approximating the implicit operator as the product of two or more factors, each of which has a much smaller bandwidth. The factors can be chosen to be upper and lower triangular matrices (in which case the scheme is called an approximate LU

factorization) or so that each of the factors contains differences in only one coordinate direction [in which case the method is called an alternating direction implicit (ADI) method]. For the LU methods, solution of the equations is achieved by marching through the field for each factor, sequentially inverting the blocks. For the ADI schemes, a block tridiagonal system of equations must be solved for each factor along each line of the mesh.

Flux computations. The above sections have represented the fluxes (i.e., the approximations to the spatial derivatives) in terms of symmetric, or central, difference approximations, and this approach, with carefully tailored artificial dissipative terms, has been very successful. Great strides have also been made using asymmetric, or upwind, differencing techniques as well. It can be argued that these upwind methods are more consistent with the physics and mathematics of the equations. Each of the eigenvalues of the Jacobian matrix $\{\mathbf{a}\}$ in Eq. (32) can be identified with the propagation of a wave, and the sign of the eigenvalue represents the direction of propagation of the wave, so the "upwind" direction for the various characteristic variables can be different, at least when the flow is subsonic. In this case, the flux vector must be "split" into two parts, each representing waves that propagate in one of the two directions, and each component is differenced accordingly. Higher order versions of a method, originally introduced by Godunov, also have been developed, in which a discontinuity is allowed between the two states separated by the mesh surface; such methods are particularly effective at capturing shock waves with no overshoot and a minimum of smearing.

Practical computations. Methods based on the solution of the Euler equations have been applied to very complex three-dimensional problems, including complete aircraft. These methods are widely used in the preliminary design stages for transport aircraft in the cruise condition, where very little flow separation is likely to occur.

4. Navier-Stokes Equation Methods

Laminar flows. As described in Section II, the relative importance of viscous effects is characterized by the value of the Reynolds number. For the large values of aerodynamic interest, the Reynolds number characterizes the thickness of the boundary layers when the flow remains attached and also determines whether the flow in these layers remains laminar or becomes turbulent. If the Reynolds number is not too large (typically, less than 10^6 , or so), the flow in the boundary layer remains smooth, and adjacent layers (or laminae) of fluid slide smoothly past one another. When this is the case, solution of the

Navier-Stokes equations is not too much more difficult, in principle, than solution of the Euler equations of inviscid flow. Greater resolution is required for the large gradients in the boundary layers, so many more mesh cells may be necessary to achieve adequate accuracy. In most of the flow field, however, the flow behaves as if it were nearly inviscid, and the methods developed for the Euler equations are appropriate. The equations must, of course, be properly modified to include the additional terms resulting from the viscous stresses, and care must be taken to ensure that any artificial dissipation added is sufficiently small relative to the physical dissipation. The solution of the Navier-Stokes equations for laminar flows, then, is somewhat more costly in terms of computer resources, but not significantly more difficult from an algorithmic point of view than solution of the Euler equations. Unfortunately, most aerodynamic flows of practical interest occur at such large Reynolds numbers that the flow in the boundary layers becomes turbulent.

Turbulence models. When the Reynolds number exceeds a critical (transition) value, then the flow in the boundary layer develops unsteady fluctuations and the flow in the boundary layer is said to become turbulent. The flow properties in turbulent boundary layers, such as the time-averaged values of velocity and shear stresses, are dominated by the effects of the fluctuating velocities and pressures. While these fluctuations still are described by the Navier-Stokes equations, the range of length and time scales over which significant fluctuations occur is so large that to resolve them all would require many orders of magnitude more storage and computer time than will be available for the foreseeable future. (Several estimates for these requirements were presented in Section II.)

The only alternative for practical calculations of turbulent flows at the present time is to make additional approximations and to introduce models for at least some features of the turbulence. The greatest reduction in computational effort is achieved by modeling all aspects of the turbulence. This usually is done by introducing the RANS equations. This approach is based upon the assumption that the effects of the turbulence can be described statistically. The solution is decomposed into time-averaged and fluctuating components; for example, the velocity components might be given by

$$u = U + u' \quad \text{and} \quad v = V + v', \quad (34)$$

where U and V are the average values of u and v , taken over a time interval that is long compared to the turbulence time scales but short compared to the time scales of any nonturbulent unsteadiness in the flow field. If $\langle u \rangle$ denotes such a time average of the u component of velocity, then, e.g.,

$$\langle u \rangle = U.$$

If a decomposition of the form of Eq. (34) for each of the flow variables is substituted into the Navier-Stokes equations, and they are then averaged as described above, the resulting equations describe the evolution of the mean flow variables (such as U , V , etc.). These equations are nearly identical to the original Navier-Stokes equations written for the mean flow variables because terms proportional to $\langle u' \rangle$, $\langle v' \rangle$, etc. are identically zero. Because of the nonlinearity of the equations, however, several terms that involve the fluctuating variables remain. In particular, terms proportional to $\langle \rho u' v' \rangle$ appear. Dimensionally, this term is equivalent to a stress, and it is, in fact, called a Reynolds stress term. Physically, the Reynolds stresses are the turbulent counterpart of the (viscous) molecular stresses, and they appear as a result of the transport of fluid momentum by the turbulent fluctuations. The implication of the appearance of such terms in the equations describing the mean flow is that the mean flow cannot be determined without some knowledge of the effects of these fluctuating components.

The Reynolds stresses must be related to the mean flow at some level of approximation using a phenomenological model. In other words, it is necessary to *model* the effects of the turbulence in order to compute the mean flow. The simplest procedure is to relate the Reynolds stresses to the mean flow properties directly, but since the turbulence that is responsible for these stresses is a function not only of the local mean flow properties but of the flow history as well, such an approximation cannot have broad generality. A more general procedure is to develop differential equations for the Reynolds stresses themselves. This can be done by taking higher order moments of the Navier-Stokes equations. For example, if the x -momentum equation is multiplied by v and then averaged, the result will be an equation describing the evolution of the Reynolds stress $\langle \rho u' v' \rangle$. Again, because of the nonlinearity of the equations, however, yet higher moments of the fluctuating variables (e.g., terms proportional to $\langle \rho u'^2 v' \rangle$ and $\langle \rho u' v'^2 \rangle$) will appear in this equation. This is an example of the problem of *closure* of the method, represented by the fact that these third-order correlations of the fluctuating variables must be known in order to solve the equations for the Reynolds stresses. It is hoped that these higher order correlations can be modeled in a more universal fashion, but there has been no clear demonstration to date that this is the case.

A variety of turbulence models has been developed to deal with this inability to close the equation set. The simplest models, based on the original mixing length hypothesis of Prandtl, relate the Reynolds stresses to the local properties of the mean flow. Such *algebraic* models are unable to account for the history of the developing turbulence,

however, and more complete models include additional partial differential equations for various properties of the turbulence. The most commonly used such models use additional equations for the turbulence kinetic energy k and the dissipation rate ε of the turbulence (or some other quantity, such as a turbulence frequency ω that determines a time scale).

Finally, even using turbulence models within the context of RANS, it is beyond the capabilities of current computers to resolve all the viscous terms in the equations for most practical problems. Since it is the gradients normal to the surface that are most important (at least for nonseparated flows), the grid spacing usually is adjusted so that only the gradients in that direction are resolved. This leads naturally to a further simplification of the equations in which the viscous terms corresponding to derivatives along the body surface are neglected altogether. Since the neglect of these terms changes the nature of the equations, in a manner similar to the boundary layer approximation, the resulting equations are called the thin-layer (or parabolized) Navier-Stokes equations.

Large-eddy simulations. The difficulty of developing generally applicable phenomenological models on the one hand, and the enormous computational resources required to resolve all the scales of the turbulent motions on the other, have led to the development of large-eddy simulation (LES) techniques. In this approach, the largest length and time scales—those that are most likely to be problem dependent—are fully resolved, but the smaller (subgrid) scales are modeled. A filtering technique is applied to the Navier-Stokes equations which results in equations having a form that is similar to the original equations, but with additional terms representing the effects of a subgrid-scale tensor on the larger scales. LES techniques date back to the pioneering work of Smagorinsky, who developed an eddy-viscosity subgrid model for use in meteorological problems. The model turned out to be too dissipative for large-scale meteorological problems in which large-scale, predominantly two-dimensional motions are affected by three-dimensional turbulence, but generalizations of Smagorinsky's model are beginning to find application in engineering problems.

Practical computations. Design computations involving the Navier-Stokes equations commonly use algebraic models for the turbulence properties, but it is an active area of research to couple the partial differential equation models to the RANS equations efficiently. The task is made difficult by the additional stiffness of the turbulence equations. The development of LES and DNS techniques remain an active research topic.

IV. CONCLUDING REMARKS

This article has described some of the ideas underlying the development of computational aerodynamics up to the present time. The emphasis has been on methods useful for determining the forces acting upon flight vehicles under relatively conventional conditions. In this situation, with the exception of the prediction of turbulent fluctuations and their effects, the physics of the processes to be predicted are well understood, and it is primarily a problem of solving known equations for practical geometries with sufficient accuracy. Great progress has been made in recent years, especially for problems for which inviscid flow models are adequate. At the present time, numerical techniques for solving better approximations to the Navier-Stokes equations are also under intense development and are beginning to see design use. These include inviscid models that allow for rotational effects (such as strong shock waves and free vortices) and the RANS equations, using a variety of turbulence models. The practical use of these methods will expand with the further development of more efficient algorithms and the continued availability of faster and more powerful computers.

Although the emphasis of this article has been on the development of engineering tools, numerical solutions of fluid mechanical problems also hold the promise of supplying fundamental knowledge in such areas as the nature and behavior of turbulence itself. Although current computers are only marginally up to the task, it should be possible in the not too distant future to compute the turbulent fluctuations themselves, at least for relatively simple flows. In fact, this is an area of intense activity, with the results of these fluid mechanics "experiments" providing detailed information under more carefully controlled situations than any physical laboratory experiment can hope to achieve, although at more limited values of the Reynolds number. The insight provided by the results of such calculations should be of great value in the development of more realistic models of turbulence for use in computations of practical interest.

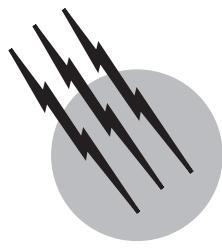
SEE ALSO THE FOLLOWING ARTICLES

AIRCRAFT AERODYNAMIC BOUNDARY LAYERS • AIRCRAFT PERFORMANCE AND DESIGN • FLIGHT (AERODYNAMICS) • FLUID DYNAMICS • LINEAR SYSTEMS OF EQUATIONS • STRUCTURAL ANALYSIS, AEROSPACE

BIBLIOGRAPHY

Agarwal, R. K. (1999). "Computational fluid dynamics of whole-body aircraft," *Ann. Rev. Fluid Mech.* **31**, 125–169.

- Anderson, D. A., Tannehill, J. C., and Pletcher, R. H. (1997). "Computational Fluid Mechanics and Heat Transfer," 2nd ed., Taylor & Francis, Washington, DC.
- Caughey, D. A., and Hafez, M. M., eds. (1994). "Frontiers of Computational Fluid Dynamics—1994," Wiley, Chichester.
- Caughey, D. A., and Hafez, M. M., eds. (1998). "Frontiers of Computational Fluid Dynamics—1998," World Scientific, Singapore.
- Chapman, D. R. (1979). "Computational aerodynamics: Review and outlook," *AIAA J.* **17**, 1293–1313.
- Hirsch, C. (1988). "Numerical Computation of Internal and External Flows," Vol. 1 and 2, Wiley, New York.
- Jameson, A. (1989). "Computational aerodynamics for aircraft design," *Science* **245**, 361–371.
- Laney, C. B. (1998). "Computational Gasdynamics," Cambridge Univ. Press, Cambridge.
- Moin, P., and Kim, J. (1997). "Tackling turbulence with supercomputers," *Sci. Am. January*.
- Moin, P., and Mahesh, K. (1998). "Direct numerical simulation: A tool in turbulence research," *Annu. Rev. Fluid Mech.* **30**, 539–578.
- Roe, P. (1986). "Characteristic-based schemes for the Euler equations," *Ann. Rev. Fluid Mech.* **18**, 337–365.



Airplanes, Light

Darrol Stinton

The Royal Aeronautical Society

- I. Introduction
- II. Categories of Light Airplanes
- III. Future Developments

GLOSSARY

Aerodynamic forces Lift (resolved normal to the direction of flight), which supports an airplane, and drag (resolved along the direction of flight), which retards its motion and is opposed by thrust from the propulsion system (and aerodynamic suction around leading surfaces of the airframe). The ability to cruise and glide long distances depends on a high ratio of lift to drag.

Aerodynamics Scientific study of the way in which air moves and reacts of surfaces to its motions.

Airflow Motion of air relative to a surface. A local airflow is tangential to a surface at some point. The “free-stream” airflow is tangential to the flight path. If the local airflow is slower than the local speed of sound, it is said to be subcritical, because the air can adjust to a disturbance. When the local flow is faster than sound, however, the molecules within it do not adjust peacefully, and the flow is called supercritical.

Airfoil (section) Surface shaped to produce more lift than drag. Wings, foreplanes, canards, rearplanes, and tails are all airfoils. An airfoil section is the profile of an airfoil when sliced from leading to trailing edge.

Airplane (aeroplane, from Gk: $\alpha\epsilon\rho\sigma\pi\lambda\alpha\nu\sigma$, a wanderer in the air) a mechanically propelled, heavier than air flying machine, with supporting surfaces fixed for flight.

Airworthiness requirements A set of standards or published code of requirements against which airworthiness can be assessed (*worthy* corresponds to the worth of, e.g., appropriateness, suitability, or fitness for use in flight). Originally, two airworthiness codes were used throughout the world, as standards set by the Federal Aviation Administration (FAA) in the U.S. and the Air Registration Board (ARB) in the U.K., which became the Civil Aviation Authority (CAA) early in the 1970s. These were joined subsequently by the complex European Joint Aviation Authorities (JAA), of which the CAA is now part. Authorities elsewhere, in Canada and Australia, for example, take national initiatives with consequences on a world scale which are of utmost importance for operators and manufacturers of engines and airframes and which cannot be ignored by the JAA, CAA, and the FAA.

Aviation (general) That part of national aviation other than the military which carries neither passengers nor freight for hire and reward.

Canard Airfoil surface ahead of the main wing. The name is often given to a tail-first airplane, because the lifting surface ahead of the wing acts in the same way as the lift-generating beak, head, and neck of a duck (Fr. *canard*).

Certificate of airworthiness Certificate issued by a certifying authority of a sovereign state to show that an

airplane meets the published code of airworthiness requirements to which it has been designed and constructed.

Coefficients of pressure, lift, and drag An airplane, pushing through air, gathers up particles of air, which are initially at rest, and accelerates them to a certain extent in the direction of flight. Those that impact on the nose and all leading edges react with a force per unit area called the dynamic pressure, a measure of the kinetic energy imparted to the air by the airplane. Dynamic pressure is a useful yardstick against which to measure other pressures on the skin, because at each speed it has a unique value. Thus, if we divide the suction or compression pressure at any point by the dynamic pressure, we have a pressure coefficient. This is negative where there is suction and positive where there is compression, and it is nondimensional. One can turn lift and drag forces into average pressures by dividing them by the wing area. By dividing through by the relevant value of the dynamic pressure, one can make these lift and drag pressures uniquely nondimensional. Hence,

$$\text{lift coefficient} = \frac{(\text{lift} \div \text{wing area})}{\text{dynamic pressure}}$$

$$\text{drag coefficient} = \frac{(\text{drag} \div \text{wing area})}{\text{dynamic pressure}}$$

Control surfaces Usually, flaplike devices arranged to move up and down at the trailing edge of an airfoil and operated by the pilot. They increase lift in a direction opposite to the way in which they are deflected. Such surfaces include ailerons (on wings), controlling roll; elevators (on tailplane or canard), controlling pitch; and rudder (or fin), controlling yaw. Trimming tabs are similar but smaller surfaces at the trailing edges of main control surfaces to remove (as far as possible) out-of-trim control forces felt by the pilot.

Engine (power plant) Prime mover installed to propel an airplane by means of thrust generated by a propeller, a ducted fan, or a pure jet.

Fuselage Payload-carrying body of an airplane (Fig. 1).

Incipient spin: Region between spin entry and a developed spin in which the outcome of the struggle between growing and changing pro-spin and anti-spin moments has not been resolved.

Spin Self-sustaining (autorotational) spiral motion of an airplane about a vertical axis, during which the mean angle of attack of the wings is beyond the stall. A spin follows departures in roll, yaw, and pitch from the condition of balanced flight. The developed spin is achieved when there is general (sometimes oscillatory) equilibrium between the predominantly pro-spin

moment due to the wings and the generally anti-spin moments due to other parts of the aircraft.

Structure Arrangement of load-carrying material (mainly within the airplane skin but often one with a modern skin) that maintains the shape and strength of the airplane under load.

Supercharger Form of pump (also called a turbosupercharger) that compresses the air supplied to the cylinders of a piston engine, and to be mixed with fuel, so as to maintain power to higher altitudes than an engine that is normally aspirated.

Units Within the world of aviation there are three sets of units. First is the old and practical Anglo-Saxon farmer's duo-decimal measures of pounds, ounces, yards, feet, inches, and miles, widely used in the world aviation markets of the U.S., which has its origins in the corresponding British Imperial system, U.S. and British gallons are an exception: 5 gallons (Imp.) = 6 gallons (U.S.), while approximately 1 gallon (U.S.) = 5 French centiliter wine bottles, and 6 bottles = 1 gallon (Imp.). The "farmer's system" has the advantage that, when operating in the field without any other aids, the average human body can be used as the module. Elsewhere, since the late 18th century (following its adoption by the French National Assembly in 1791 and the conquests of Napoleon Bonaparte), there followed what might loosely now be called "market metric," in which the kilogram became the unit of weight and the meter the unit of length. The English yard pace of 36 inches (equivalent to the medieval clothyard arrow, fitted to a fully drawn bow) is shorter than the meter, which, at 39.37 inches, is equal to the girth of a well-fed Frenchman. In the scientific community, the metric system was rationalized in the 1960s by the academically dignified *Système International d'Unités*, or SI, in which the kilogram became the unit of mass, and the Newton the unit of force (roughly equal to the weight on the hand of one small apple).

LIGHT AIRPLANES are winged flying machines extending in weight from microlights and ultralights (the weights of which are subject to national differences) to the upper limit of 5700 kg (12,500 lb), which is the maximum for operation by a single pilot. Most light airplanes are propelled by a single piston engine, up to a weight of 2730 kg (6000 lb), although increasing numbers intended for air taxi and business use are fitted with turbopropeller units. Beyond 6000 lb, light airplanes are increasingly twin-engined, although the light weight and high power of turboprop engines make it possible to replace twin piston engines with a single turbine-propeller, up to the limit for single-pilot operation.

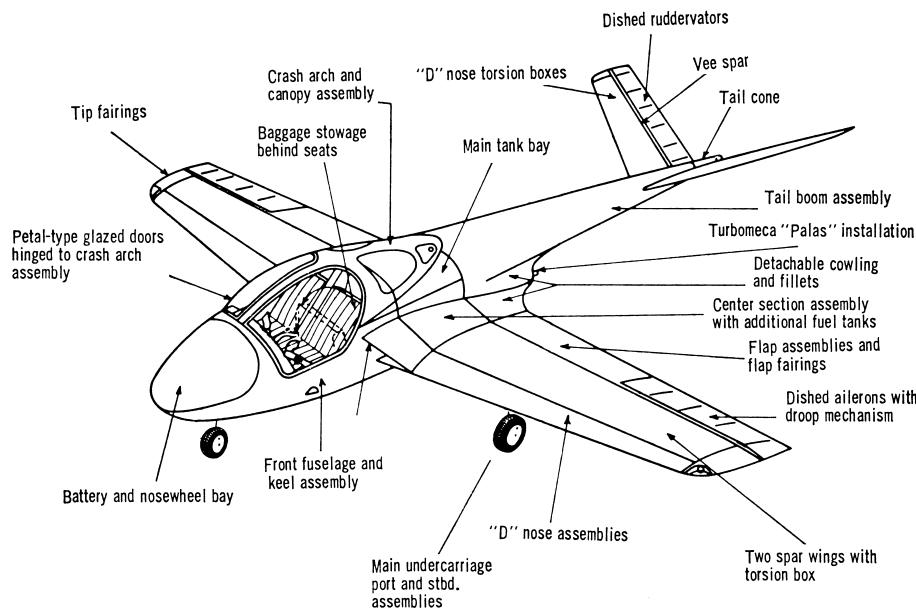


FIGURE 1 Main components of a jet-engine, two-seat light airplane.

Inevitably turbojet-powered light airplanes are being seriously considered as a possibility for the future. This presents a problem for the regulatory authorities because of their potent performance. They are able to climb above the weather, fly fast, and penetrate controlled airspace. Pilots are needed with higher skills, training, and medical categories than are normally expected for nonprofessional (private) pilots flying conventional light airplanes. They also need appropriately advanced levels of radio and navigational equipment in the event of emergencies that may cause them to divert to major airfields because they are unable to fit into average airstrips.

Although the trend is toward high performance and propulsion by gas turbine engines—turbopropeller and turbojet (Fig. 1 shows typical features of a light, two-seat jet airplane)—there is an enthusiastic trend in the opposite direction toward lightness, simplicity, and flying purely for pleasure. The microlight and ultralight airplane, which started as single seaters, are constantly being improved to carry two persons, in the face of tight restrictions upon weight imposed by the certifying authorities.

I. INTRODUCTION

About four-fifths of all general aviation (GA) aircraft flying worldwide are light airplanes. The total has remained steady for many years, in spite of widespread international recession in the light airplane industries through the 1980s and 1990s. Since Wilbur and Orville Wright flew on December 17, 1903, at Kittyhawk, North Carolina, the

bulk of all flying machines have been light airplanes. The most popular spectator sport after football is the airshow. The most popular aircraft are generally light antique, vintage, and classic, and many “warbirds” (ex-military) in refurbished, rebuilt, and replica form are highly prized. Operation of such aircraft is a major industry. Their standards are set and managed by *The Confederate Airforce* and the *Experimental Aircraft Association (EAA)*, *Warbirds Division*, in the U.S., and by the *Historic Aircraft Association (HAA)* in Britain.

The Second World War started in 1939 with the British *Supermarine Spitfire Mk 1* (Fig. 2) opposing the German *Messerschmitt Bf 109E* (Fig. 3). Both weighed less than 2730 kg (6000 lb) and deserve to be rated as two of the most potent light airplanes ever built. Willy Messerschmitt said of his 109: “... I made an attempt to equip the smallest possible light aircraft with a powerful engine in order to produce a fighter which exceeded in its performance anything that had been seen before.” Today, they are not only representative warbirds, but in England there is a production line of various marks of Spitfire built by groups of enthusiasts. Until the advent of the gas-turbine or jet engine, the bulk of all significant airframe and engine research and development necessary to expand the boundaries of flight was carried out by light airplanes. Such areas included handling at very low airspeeds (stall and spin investigation); flight at high altitude; high-speed flight (the Schneider Trophy for high-speed seaplanes was won outright in 1931 by the *Supermarine S 6B*, weighing only 2760 kg [6086 lb], while the previous winners, the *Supermarine S5* and *S6*, weighed much less than 2730 kg);

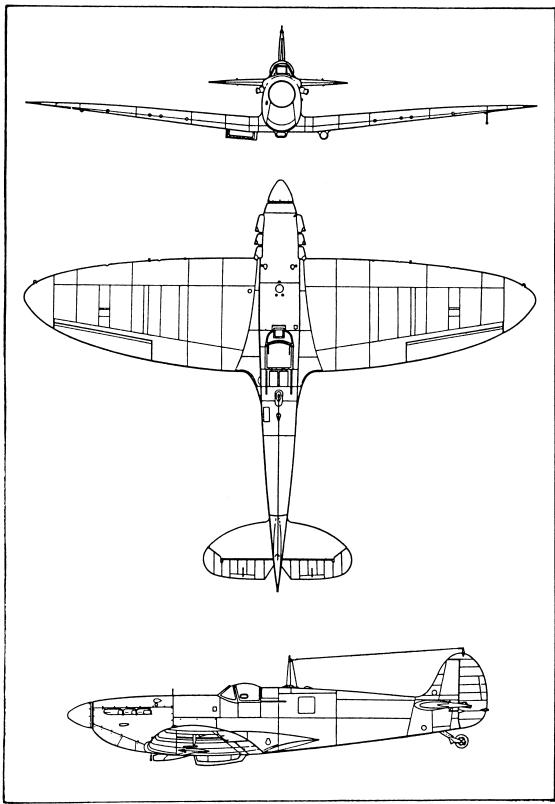


FIGURE 2 Supermarine Spitfire Mk 1, 2629 kg (5784 lb), 355 mph at 19,000 ft, ceiling 34,000 ft, 8 × 0.303 guns. Today this Spitfire would be rated as a warbird 2 (United States) or “military and vintage” (United Kingdom) light airplane. [Courtesy of Derek Wood and Derek Dempster.]

and most long-range record-breaking flights. In short, light airplanes have spearheaded aeronautical advance most of the time. Indeed, they continue to play a leading role in a number of critical areas of aerodynamic and structural research, development, and design, especially in the use of new materials. Figures 4–6 show typical examples of light airplanes. Among them, the design emerging from the stables of Burt Rutan, Mojave, CA, have broken fresh ground as dramatically as that of the diminutive *Pou-du-Ciel* (*Flying Flea*) designed by Henri Mignet of France in the 1930s (Fig. 7), the great grandfather of all really light, home-built airplanes.

General aviation accounts for more than 96% of the Western world’s civil airplanes, operated by more than one million private pilots. GA movements exceed by between three and four times those of the air transport sector. In spite of recession and a down-turn in sales of light aircraft through the 1980s until the late 1990s, the period was marked by manufacturers investigating future development of light single-turboprop airplanes, in spite of prices increased by anything from 10–90% over those of roughly

comparable piston-engined types. The market in the new Millennium is that of the wealthy businessman who needs high performance in a time-driven commercial world.

Economic and environmental factors are of critical importance. Modern general aviation airplanes fall into two broad categories: those intended to have the commercial efficiency of the high-performance, high-value private car and those which are aerobatic, constructed for the sportsman-pilot. In fact, when compared with light airplanes built between the two world wars, in the 1920s and 1930s, modern machines are less efficient when assessed in terms of Everling Number, which is the ratio of propeller efficiency divided by the overall drag coefficient of the aircraft.

General aviation airplanes, like larger commercial aircraft, must also satisfy operational needs in the safest and most productive ways for those they are designed to serve.

$$\text{productivity} = \text{weight of payload} \times \text{block speed}$$

where block speed is the distance flown in the time from brakes “off” to brakes “on” again at the end of the flight.

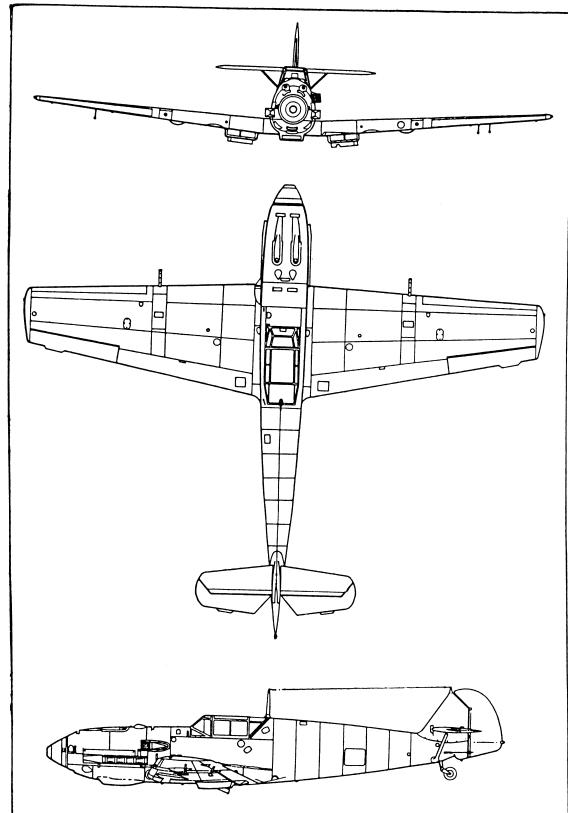


FIGURE 3 Messerschmitt Bf 109E-3, 2505 kg (5523 lb), 354 mph at 12,463 ft, ceiling 36,091 ft, 4 × MG 17s, and an engine-mounted MG FF/M cannon. Another warbird or “military and vintage” light airplane by the standards of today. [Courtesy of Derek Wood and Derek Dempster.]



FIGURE 4 Piper Malibu.

The concept of productivity falls into two parts: weight of payload carried and speed. Carrying a large weight of payload (the revenue-earning load) requires volume—which implies ease of loading and unloading—and low structure weight. More accurately, the empty weight of the airplane must be as low as is efficient. This means that structure, power plant, system, and equipment weights must also be as low as possible, within the rational limits set by airworthiness requirements.

The achievement of high block speed implies high cruising speed in flight: as fast as possible for the weight of fuel

consumed. There is also, by implication, the need for easy turnaround on the ground, which leads back again to layout and structural design.

A useful way of drawing these various threads together is by means of the Breguet range equation devised by Louis Breguet, pioneer French aircraft designer (1880–1955). The equation brings together three discrete items, which enable one to calculate the range of an airplane for a given payload. In fact, one should not consider payload in isolation, but disposable load, where:

$$\text{disposable load} = \text{payload} + \text{weight of fuel}$$



FIGURE 5 Italian Partenavia P68 (Partenavia Costruzione Aeronautiche SpA).



FIGURE 6 *Lazair* microlight.

Fuel and payload are therefore mutually exchangeable. Either a large payload may be carried for a short distance, or a small payload may be carried for a much longer range.

Three terms together show how far an airplane might fly. The first is

$$\frac{\text{propeller efficiency}}{\text{specific fuel consumption}}$$

for piston and turboprop engines, or

$$\frac{\text{cruising speed}}{\text{specific fuel consumption}}$$

for a jet. Here, specific fuel consumption is measured as the weight of fuel flowing divided by the net power of a



FIGURE 7 The French *Le Pou du Ciel* (*Flying Flea*), designed by Henri Mignet in the early 1930s, took the aeronautical world by storm and remains the progenitor of all popular home-built projects. Three design faults led to the original being banned. Once the faults were rectified, later modified variants continued to fly and are still among the safest handling and least complicated of all light airplanes. [Courtesy of former Sampson Low, Marston & Co., Ltd., circa 1934–5.]

piston (or turboprop) engine, or the net thrust of a jet. It is the usual indicator of merit of an engine. The second is

$$\frac{\text{lift}}{\text{drag}}$$

a measure of aerodynamic efficiency. The third term is

$$\frac{\text{weight of fuel available for cruising}}{\text{weight of the airplane}}$$

which is an indication of how light is the remaining weight of structure, power plant(s), systems, equipment, and operating crew. Range is obtained numerically by multiplying the three terms together,

$$\text{propulsive} \times \text{aerodynamic} \times \text{structural}$$

while introducing appropriate constants and using compatible units. They are discussed in Section III.

A. Regulation of Air Safety

Two global systems regulate safety in the air: one American, the other European. That of the United States, the Code of Federal Regulations, is a codification of the general and permanent rules published in the Federal Register by the executive departments and agencies of the federal government. Administration of the Federal Aviation Regulations (FARs) is by the Federal Aviation Administration (FAA), Department of Transportation (DOT).

In Europe, the British Civil Aviation Authority (CAA) has administered the dominant system of regulations, made under powers contained in the Civil Aviation Act, passed by Parliament. A primary task of the CAA has been the certification of aircraft to the code published in the *British Civil Airworthiness Requirements* (BCARs). However, great changes are taking place within Europe which have led to revision of the scheme of things. Most European countries have grouped together as the *Joint Aviation Authorities* (JAAs), which is publishing and administering the European joint airworthiness code, contained in the *Joint Aviation Requirements* (JARs).

There are two almost identical global codes: FARs and JARs. It is argued that there is merit in having two, because they are more likely to maintain the cutting edge of standards than if there were only one. By the time that JARs have been adopted completely, it is possible (but by no means certain) that the JAAs will have been renamed the European Aviation Authorities.

Regulation of safety in the United States and United Kingdom has a number of elements that fall broadly under the following common headings:

$$\begin{aligned} &\text{fiscal constraints} + \text{enforcement constraints} \\ &= \text{publicly acceptable level of safety} \end{aligned}$$

In the United States, the equation can be written as:

- law enforcement
- + insurance regulations
- + ground inspection by, or required by, FAA
- + ongoing education by FAA
- + positive monitoring by FAA
- = publicly acceptable level of safety

Contravention of any of the terms or failure to comply is penalized financially either directly or indirectly by large court and attorney's fees, increased insurance premiums, products liability, error and omission (or other relevant insurances), and loss of inspection approval by the FAA of mechanics designated to carry out certificate of airworthiness inspections on behalf of the administration, an accolade much sought after and highly prized.

In the United Kingdom, there is no precise equivalent of ongoing education. Instead, the Civil Aviation Authority (CAA) equation might be written as:

- operational regulation and inspection
- + airworthiness ground inspection
- + monitoring + flight test
- + law enforcement
- + insurance attitudes hardening along U.S. lines
- = publicly acceptable level of safety

The airworthiness terms have been grouped together because a flight test by the CAA in the U.K. is a form of airborne inspection.

Organization involved with management, good order and discipline of GA, which contribute toward achieving a publicly acceptable level of safety for those tasks, roles, and aircraft coming under their aegis, include the following:

- AAA: U.S. Aerial Applicator's Association (agricultural and related aviation)
- AOPA: U.S. Aircraft Owners and Pilots Association (the private sector)
- EAA: Experimental Aircraft Association (amateur builders, but including owners and operators of a wide range of vintage, classic, and exmilitary (warbird) aircraft)
- GAMA: U.S. General Aircraft Manufacturers' Association (construction and fabrication)
- GAMTA: British equivalent of GAMA, including traders
- NBAA: U.S. National Business Aircraft Association (corporate operators)

On the side of the government, the following might be added:

- NASA: National Aeronautics and Space Administration
(which runs research and test programs)
- NTSB: U.S. National Transport Safety Board

II. CATEGORIES OF LIGHT AIRPLANES

A. Business Airplanes

The decision by management to invest in an airplane for company business rests on a simple question: Will it increase profits?

In seeking an answer, the first item of high expenditure is the cost of time and travel of its executives—to customers, distributors, suppliers, outlying factories, and subcontractors. Such flexibility and freedom as the business aircraft bestows are often of great advantage when it comes to avoiding the restricting effect of scheduled services and large concentrations of passengers at airport terminals. Many negotiations are cut short by check-in times for scheduled flights. Many decisions are cramped by delays caused by disrupted scheduled services and fighting the traffic jams to and from busy terminals.

The case for a company aircraft hangs on projected utilization by all employees, based on a survey of travel time and expenses. If the projected annual utilization rate is 300 hr, then a company aircraft might be of use.

Most business airplanes are small four- to six-seaters, with twin engines for all-weather operation. It is useless thinking of anything less. Most single-engine company aircraft are limited to daylight and remaining in visual contact with the ground. Airplanes with only one engine cannot meet the common-sense requirement that, if the engine fails, the aircraft must be able to glide to a point suitable for a safe landing. Above clouds or at night, without sight of ground, this obviously will not work.

According to a survey carried out for the NBAA, 95 of the top 100 companies operate their own aircraft. They are said to account for 84% of the top 1000 companies' sales, 86% of their profits, and 85% of their assets. They are alleged to be better in sales and 49% better in profits per employee than companies without aircraft.

B. Air Taxi Airplanes

On-demand operations with all-weather flexibility involve the use of airplanes similar to those used for business flights, carrying high-value executives. Avionics, the communications and navigational equipment carried, are comprehensive and expensive. Air taxi pilots can be expected to have an airline pilot's license and at least 1000 hr public transport operations. They are renowned for their initiative and ability to "get you there—anywhere."

C. Personal and Private Airplanes

As already noted, most personal and private aircraft have single engines. However, it is within this category that some of the most interesting developments are taking place, among home-built sport and recreational aircraft, including microlights and ultralights.

According to the FAA, a home-built or amateur-built aircraft is one for which the owner has done at least 51% of the work needed to make it fly. This was to prevent manufacturing companies from avoiding the airworthiness regulations of FAR 23 by selling partially constructed kits to the innocent and unwary. When completed, such aircraft are rated as "Experimental—Amateur Built," which has its own limitations as to use.

In addition to sport and recreational airplanes are ultralights, antique, classic, and warbirds (military and vintage in the United Kingdom).

1. Sport and Recreational Airplanes

Sport airplanes range from simple hang gliders with engines to advanced all-metal monoplanes of very high performance. The impetus at the ultralight end of the range has come from three directions, the first two being a direct consequence of space technology:

1. New manmade materials: composites of various kinds that make it possible to try much lighter structures and new forms.
2. The flexible Rogallo-type wing: a triangular or lozenge-shaped membrane wing developed by Francis M. Rogallo around 1945 and employing weight shift by the pilot to achieve control. This is most adaptable to microlights.
3. The three-axis control, which stems from the experiments of Volmer Jensen around 1941 and which has aerodynamic surfaces, like a conventional airplane, for control in pitch, roll, and yaw.

2. Microlights and Ultralights

The legal definition of a microlight differs, depending upon the regulations of the country to which it is subject. Both weight-shift and three-axis hang-gliders, when fitted with engines, are described as ultralight in the U.S. or microlight in the U.K. The name "ultralight" means something else in the U.K. and to different authorities elsewhere. Either way, the airplanes are small and very light and operate with small, high-revving, and therefore noisy, engines. In due course, quieter solutions will be found but, as this is written, noise is perhaps the most offensive aspect of engines which turn at speeds close to 5000 revolutions per minute.

Interest is now being shown in electric propulsion, which helps to alleviate the problem of noise but pushes weight (mass) upwards to challenge nationally defined legal limits.

3. Antiques

By definition, an antique airplane is one designed and built before December 31, 1945, so the category covers more than 50 years of aircraft production. Restoration is often frustrating, and owners can be led astray by self-styled experts who, with knowledge of modern methods, can be out of touch with experience, wisdom, and judgment which is older but more relevant.

4. Classics

The definition of a classic airplane is one built or designed before December 31, 1955. Such aircraft display technology proved during the Second World War, and they were built for a post-war boom that never came because the automobile was not yet ready to be replaced by the airplane.

5. Historic and Warbird (Military and Vintage) Airplanes

"Historic aircraft are those generally regarded as possessing such technical features, flying qualities, operational and historic associations as to warrant uncommon knowledge or skill when preserving them for display and use in representative working order" (Historic Aircraft Association, U.K., 1978).

The important words in the definition are "representative working order." It is ethically and historically wrong to interfere with the working of any artefact, such as an airplane or a ship, because doing so destroys the point and purpose of those tasked with carrying out future research.

Combat and training airplanes used in the Second World War represent the ultimate in piston-propeller aircraft development. They form a group of their own, although jet airplanes are now coming into private ownership.

In the United States, the Confederate Air Force, a charity that draws tax relief for its owner-members, provides a focus for warbird interests, together with the Warbirds of America organization of the EAA.

In the United Kingdom, the Historic Aircraft Association, a limited company, administers a register of pilots qualified to display such valuable and attractive aircraft. It uses the large pool of experience among its members to educate and advise on the interests of flight safety and keeping rare historic aircraft accident free.

6. Aerobatic Airplanes

Aerobatics have come to the fore as an internationally competitive sport. Maneuvers have been developed that

are far beyond anything previously conceived, and specialized airplanes are designed to fly them. Jose Luis de Aresti, a Spanish aerobatic pilot, devised a system of assessment in which aerobatic maneuvers are analyzed in families, from 1–9. Armed with the basic families it is possible to combine the maneuvers into sequences, the most demanding of which are flown in various international aerobatic contests. The Aresti system is only the foundation of contest rules, which can modify the Aresti concept considerably.

During aerobatic contests the object is to remain within a box of airspace of defined limits. This requirement results in the need for the following:

1. High structural strength
2. The ability to maneuver at comparatively low airspeeds (too much kinetic energy through too much speed makes it difficult to stay within the box)
3. Crisp responses to effective controls and the ability to fly either way up or on one's side with ease
4. Plenty of power per unit weight
5. Ample wing area for the weight of the airplane, resulting in relatively low wing loading (i.e., weight per unit area)
6. Quietness for obvious social reasons (although few airplanes achieve it)

D. Training Airplanes

Although most training airplanes are identical with aircraft bought and owned for business and other purposes, for which their pilots were trained, nevertheless there is a case for specialized machines at the *ab initio* stage, when first learning to fly. The best trainers in the past have been those that were demanding, but which could then be flown out of a difficult corner with ease by a pilot who took the correct actions.

Around two-thirds of all flying accidents are caused by pilot error. Of those, two-thirds are the result of mishandling at low airspeeds, leading to stalling and/or spinning.

A stall occurs when the airspeed is so slow that lift is no longer equal to or greater than weight. The result can be an accident when too close to the ground to recover.

A spin occurs when one wing is more stalled than the other, so that the airplane departs from its flight path in pitch, roll, and yaw. Then, it begins to autorotate in an asymmetrically stalled spiral descent. A pilot must often take a set of seemingly illogical actions with the flying controls to recover from a spin.

It is now FAA policy to train pilots to avoid stalls and spins instead of how to recover from them and to have airplanes designed that will neither stall nor spin. The CAA in Great Britain is taking the same route. Having

said that, nevertheless, there is a considerable body of experienced opinion that is uneasy. A spin, for example, is the worst thing an airplane can do to its pilot—like a horse bolting. One would not be very happy if one's child were taught to ride but not shown what to do if the horse bolted one day. In a similar way, there is a case for airplanes that can be used for training pilots in developed stalls and spins, just as easily as they can be taught to avoid them. Aircraft with such livelier qualities then make it much easier for those pilots who wish to graduate to advanced aerobatic machines that are more demanding.

E. Agricultural and Aerial Work Airplanes

The last 20 years wrought marked changes in the number of agricultural aircraft now operating around the world. A rough estimate in 1983 suggested that there were at least 28,000 engaged in crop spraying and in the application of pesticides, of which two-thirds were airplanes. But aerial activity of this kind is now severely limited, especially in the U.K. and Europe. When it is necessary to spray crops the farmers now tend to carry out the work themselves, using a tractor fitted with long spray-booms. The public is understandably nervous of about the effects of what are known to be toxic pesticides. In 1983, it was estimated that the number of airplanes engaged in the work would almost double by the millennium. Instead, their growth in number is much reduced, except in the third, developing, world.

Among aerial work airplanes, some of the most highly regarded are those modified for fire fighting. The growth in vacation and holiday traffic and tourism in places such as the coastal regions of Spain and Mediterranean Europe bring the constant threat of fires started accidentally, as well as deliberately by arson.

Many agricultural airplanes started life as something else, with piston engines, and the trend is to re-engine with costly turbopropeller-units. While this increases operational flexibility and effectiveness, it introduces problems for other certifying authorities that have different standards, having to cope with aircraft designed originally for a different task, often at lighter weights. Those particularly relevant here affect stability with different, defined power settings and configurations; flying control; trimability; and engine handling. For example, it should not be possible to pull back the power lever of a turbo-prop airplane in flight and enter the reversed pitch range inadvertently, when the propeller behaves like a big air-brake that brings the airplane almost to a stop and literally kills the airflow over the tail control surfaces. Yet, it is vital at times for a pilot of an agricultural airplane, wafting down the approach path, 55% of the takeoff weight gone, to be able to put the airplane down accurately on

a short strip, perhaps in a gusting crosswind, and to select reverse thrust instantaneously. Therefore, one cannot, perhaps, apply quite the same design approach to an agricultural airplane as to an air taxi or commuter airplane with the same engines but full of passengers' and flying into an airport.

Agricultural machines can carry big loads in hoppers capable of holding spray solutions, dust, seed, water, or even fuel. The hoppers, when filled with fuel, enable some airplanes to achieve trans-Atlantic ranges or be ferried over long distances. In the West there is a tendency to design hoppers that occupy a considerable length of an airplane. In the Russian Federation and Eastern Europe, tall, upright hoppers are favored. These have the advantage of tightly limiting movement of the center of gravity due to the free-surface effect of a large mass of liquid. The former, while resulting in lower frontal area and less obvious drag, can cause trim and longitudinal stability problems with heavy loads of liquid surging backward and forward over relatively long distances which upset the trim to a certain extent, in spite of internal baffles. Hoppers placed vertically do not tend to upset the center of gravity of the airplane quite so much.

It is among agricultural airplanes that some of the most interesting developments are to be found. New structural layouts and materials are being tried to achieve lightness compatible with the heaviest, most productive loads. New arrangements of load-carrying and lifting surfaces—wings, tails, and foreplanes—are being investigated to achieve a high lift/drag ratio and docile low-speed handling qualities. New engine arrangements based on small gas turbines are being tried. Finally, experiments are being carried out with ultralight and microlight airplanes, which may enable farmers to spray their own crops precisely with microscopic drops of pesticide without flying beyond their own fences.

III. FUTURE DEVELOPMENTS

Although it can be argued that light airplanes are eminently sensible, carrying more than one-half of the air-traveling population more than 70% of the miles to 100% of airports, they will not survive in these fuel- and ecologically conscious times if they fail to fit the mood of the public. Light airplanes have to be shown to preserve the environment, and accident rates are given a bad press. Legislation has a subtle and pervading effect on the course of development.

A. Pressures of Legislation

The prospect of a product liability lawsuit haunts the aircraft manufacturer. Product liability is a concept meaning,

in general, the legal responsibility of a seller of a product for defects that cause death, injury, or damage to its user. The seller is usually the manufacturer, but a distributor, lessor, or retailer may also risk attracting a liability. The legal responsibility is usually in the form of money damages paid by the seller to the injured person or family. Three theories are used to impose the responsibility on sellers:

1. Negligence: The failure to exercise the duty of care of a reasonably prudent manufacturer or seller in the creation and distribution of a product.
2. Breach of warranty: May include a warranty of fitness for the purpose for which a product was created and a warrant of merchantability.
3. Unreasonably dangerous: If a product is defective and the defect makes it unreasonably dangerous, causing injury or damage to someone who has neither misused nor abused it, and if the product is substantially the same as it was when it left the manufacturer, then no matter how much care has been taken to make it safe for the user, the manufacturer is still liable.

The need for safety and to make airplanes as foolproof as possible—because all airplanes bite fools—leads in four fairly clear directions of development: (1) avionics and cockpit design, (2) aerodynamics, (3) structures and materials, and (4) propulsion.

Flying is much safer than road transport, accounting for ~2.5% of the people killed in all transport accidents in the United States in 1983. Road transport accidents were 92.2% of the whole. Even so, the most prevalent causes of light airplane accidents are inadequate preflight preparation and planning, which often include insufficiently relevant continuation training, leading to pilot overload in flight when faced with circumstances that were not anticipated.

B. Avionics and Cockpit Design

An important feature of a modern light airplane that is fit to fly airways under Instrument Flight Rules (IFR) is its capacity to navigate to almost any airport in the world. Very low frequency (VLF) Omega navigation and very high frequency (VHF) satellite systems are being developed. Omega is already being used for en route navigation where 2- or 3-mile position accuracy is sufficient. Further ahead is proposed a network of satellites with time and space synchronization that will provide worldwide three-dimensional positioning, within 8–10 miles.

Thus, the relationship between pilot, cockpit, equipment, and airplane represents an integrated man-machine system. Such integration will become even more important. If proper attention is not given to the system now,

during the design stage, it is possible to overstress a pilot in an emergency, putting him beyond his limits by facing him with a more intensive workload than he should have at the time.

The microchip revolution has produced navigation and flight control equipment with which a qualified pilot can fly virtually anywhere under all conditions. These aircraft place more strain on the pilot than the customary tension level during single-pilot IFR operations, possibly with inconsistent passengers and equipment failures to spice the atmosphere. The ability of any pilot to cope depends on the level of stress arising from the workload plus a number of other factors, which are usually emotional. These, and particularly the last-named, change daily and even hourly, depending on such vagaries as blood sugar level or, more extremely, divorce or recent bereavement.

Cockpits and installed equipment must be well designed, with everything falling readily to hand. Under stress, even the most seemingly logical and obvious color-coded instrument presentation can become incomprehensible. All human beings revert to earlier learned patterns of behavior when overloaded. This is the greatest reason why manufacturers should try to settle on just one optimized instrument panel arrangement and cockpit layout. So far, they have not been successful.

C. Aerodynamics

An airplane in flight causes drag, the force tending to resist its motion. Light airplanes are subsonic almost without exception: They fly at speeds less than that of sound and do not, therefore, expend much additional energy in overcoming shock waves and other phenomena arising from compressibility. However, light airplanes that will reach high subsonic airspeeds with supercritical airflows are already promised, with a number of advanced designs now being studied. High airspeeds are expressed as Mach numbers, that is, the true airspeed divided by the speed of sound in the surrounding atmosphere, after Ernst Mach (1838–1916), the Austrian philosopher and physicist who first noticed its significance. At speeds of less than that of sound, drag is caused as follows:

1. Form, due to displacement by the bulk and bluffness of the form of the machine, as it shoulders the air aside
2. Skin friction, due to the viscosity of the air as it slides over the “wetted area” of the airframe
3. Interference with the airflow by adjacent parts of the airplane

In addition to these purely parasitic components is drag that is lift dependent or induced by the particular act of

generating lift by the wings and other surfaces causing air to form a lifting vortex system.

The art of shaping an airplane lies in achieving low form, frictional, and interference drags, by keeping the wetted area small and forms streamlined, on the one hand, while providing enough wing span, on the other, to reduce the energy lost in generating lift.

A wing (or any other similar surface) generates a lift force normal to the discretion of flight by pushing air downward at a given rate. The “downward” acceleration given to a mass of air, involving a change in momentum, produces an “upward” reaction, felt as lift. The size of the mass affected depends on the span of a wing and its true airspeed. The longer the wing span the smaller the “downward” acceleration and kinetic energy needed. It follows, of course, that if a long-spanned wing does less work in imparting a downward kinetic energy to the air than one that is short, it also needs less fuel to keep generating the required lift at the same airspeed.

As a general rule for least drag, thickness and depth of airfoil sections normal to the direction of flight must be small compared with chordwise length tangential to the airflow. But this leads to heavier structures and can negate aerodynamic gain from having a long wing span. Designers must therefore optimize their wings so as to avoid weight penalties on payload and fuel carried.

1. Boundary Layer Airflows

Much work has been done on making the boundary layer flow of air next to the skin of an airplane laminar, that is, one that slides past in thin, smooth layers. The thinner and smoother the boundary layer, the lower the skin friction drag. Roughness causes a laminar boundary layer to “trip” into a thicker turbulent form, which absorbs more energy and slows the airplane by increasing frictional drag. At low speeds and when splattered with rain drops or bugs, laminar flows break down with comparative ease, decreasing lift and increasing drag. Some very clean laminar flow control surfaces become ineffective in rain. Some motor gliders with laminar flow wings fail to get airborne in the field length available because of raindrops on smooth surfaces. Paradoxically, a *moderately* turbulent boundary layer, while causing more drag than one that is laminar, sticks better to the airframe surface, refusing to break away in the large lumps of turbulence that follow the uncontrolled breakdown and separation of a laminar boundary layer.

Much work is done in the wind tunnels of NASA and the industry to shape new low-drag, high-lift wing sections. An airfoil section is especially shaped so as to make particles of air *circulate* around curved paths, while being displaced downward, as shown in Fig. 8. The greater the

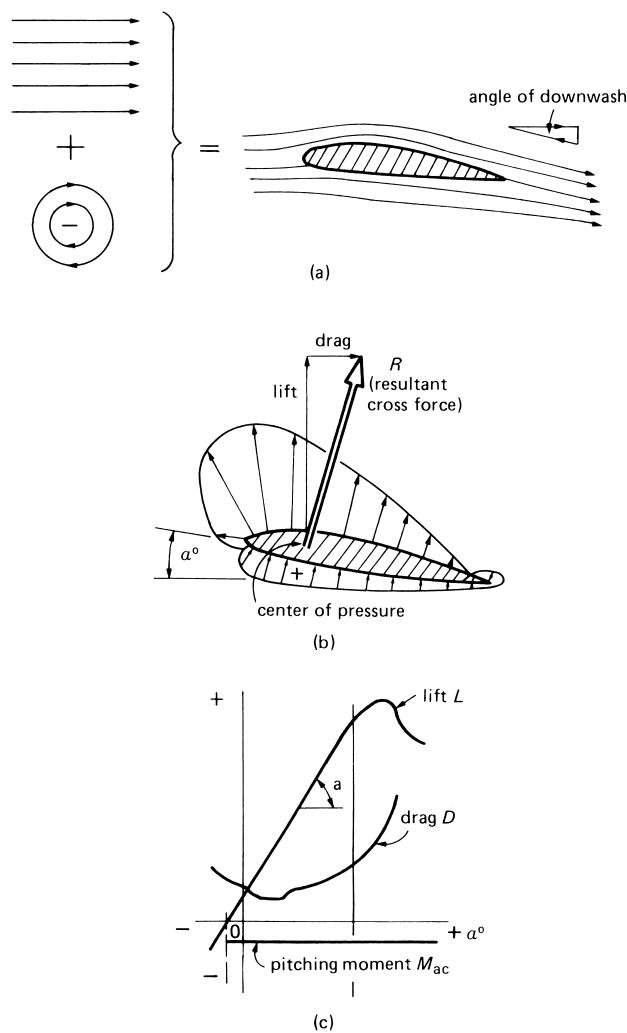


FIGURE 8 Circulation as the source of lift (and drag) generated by an airfoil surface. (a) When a circulatory (bound vortex) motion is induced in a straight, steady flow by an airfoil, both motions can be added vectorially to produce the flow pattern seen around a lifting surface in a wind tunnel. Note the way in which the flow leaves the trailing edge with a downwash component. (b) The flow pattern changes with angle of attack, suction over the top, and compression underneath, the lobes being typically as shown within the normal working range of lifting angles of attack, say, approximately -2° to $+12^\circ$. (c) The circulatory motion results in an aerodynamic cross-force, which can be resolved into lift L , normal to, and drag D , parallel with the flight path. There is also a nose-down (negative) pitching moment, which has a more or less constant value when measured about the aerodynamic center of the airfoil. [From Stinton, D. (1983). "The Design of the Aeroplane," Van Nostrand-Reinhold, Princeton, NJ.]

net circulation imparted to the air as it sweeps past, the stronger the lift.

Airfoil sections are designed for efficient cruising, docile high-speed characteristics, and ease of handling at low airspeeds. For high speed, one needs thin wing sections to reduce the displacements and accelerations im-

parted to the air in passing. The more vigorously the air is disturbed and made to follow a lift-generating path, the more energy it takes to do so.

It follows that, if an airplane flies fast, the air is forced over the skin at even higher speeds in places, and if the speed of flight is more than ~ 0.6 times the local speed of sound, $M \approx 0.6$, local flows may become supersonic. After the disturbance is past, the air must then slow down again, which it does, giving up its acquired extra energy through a narrow, disruptive shock wave.

Conventional airfoil sections, designed for subsonic flight, generate intense systems of shock waves when pushed to faster speeds. Because of their inherent limitations in this respect, new *supercritical* airfoil sections have been designed. Figure 9 shows a rear-loaded airfoil section compared with a conventional section. In the United States, much research into such sections is being carried out by NASA, after the original work by Richard T. Whitcomb. It has been paralleled elsewhere (e.g., in the U.K. by the former Royal Aircraft Establishment (RAE), now the Defence Evaluation and Research Agency (DERA), and British Aerospace (BAe)). Rear-loaded sections are so called because they are designed to generate lift farther aft, over a larger part of the chord (between the leading and trailing edge). This spreads the distribution of pressure, reducing the peak associated with high local velocity in the relative airflow and so reducing the local peak velocity. A smooth, locally supersonic airflow can be maintained farther aft than with a conventional section. The final shock wave, when it comes, is weaker and less drastic in its effect of blasting the boundary layer behind it. Such blasting causes high drag and may be disruptive enough to lead to gross loss of lift. Effects of this kind, coupled with loss of stability and control, once led to a belief in a sound barrier.

Another area of focus is that of low-speed handling to reduce stall-spin accidents. Effort centers around the reduction of data obtained from wind tunnels, flight tests, and simulators.

2. Configurations

"New" configurations (arrangements) of wing, tail, and body surfaces have been tried in recent years, with varying degrees of success. The conventional layout (wing in front, tail at the back) has been challenged, no matter how transiently, by airplanes with a smaller horizontal "tail" ahead of the wing. There are some advantages, such that the lift of both foreplane and rearplane (or wing) surfaces generates lift in the right direction, whereas when the tail is at the back, tail lift is often downwards, for trimming. Tail-first aircraft can also be built with shorter bodies which have less wetted area and, in theory, less drag (Fig. 10).

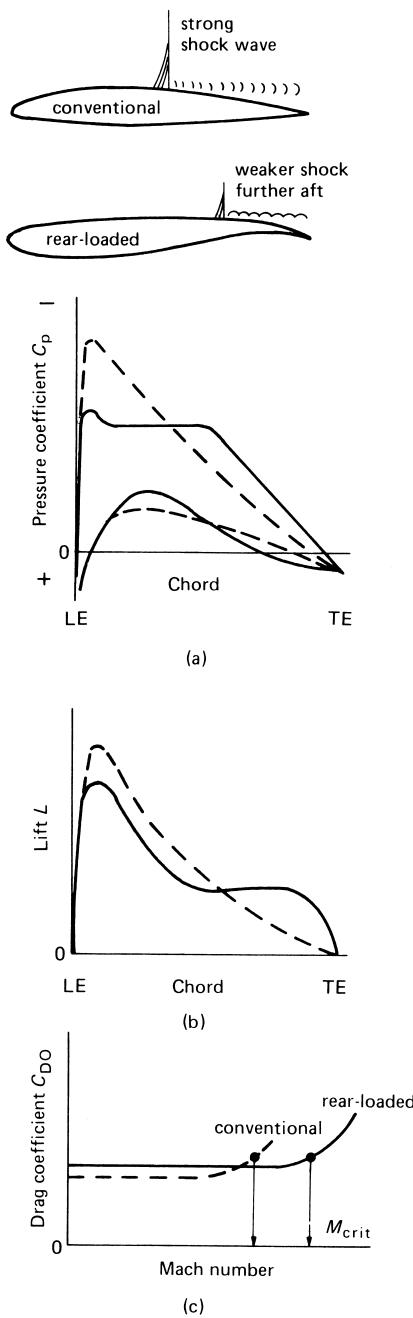


FIGURE 9 Comparison between conventional (---) and rear-loaded (—) airfoil sections. (a) Pressure distribution; (b) lift distribution; (c) improved critical Mach number. See Glossary for pressure and drag coefficients. [From Stinton, D. (1983). "The Design of the Aeroplane," Van Nostrand-Reinhold, Princeton, NJ.]

Disadvantages are considerable. Foreplanes are destabilizing and must be more highly loaded than the main wings to force them to stall first, so pitching the aircraft safely nose down, to regain flying speed (to pitch nose up at the stall is to make matters dangerously worse). It is



FIGURE 10 Canard airplane: the Rutan Model 40, *Defiant* (Howard Levy).

difficult to fit flaps for slow flight. They cause large out-of-balance moments, unless fitted front and rear, which is complicated and expensive. Therefore, to tailor a tail-first, or canard, design for operations from defined airfields, more wing area is needed to achieve slow enough touch-down speeds. And such aircraft are less maneuverable and agile—unless, like high-performance fighters, they are artificially controlled and stabilized by complicated electronic systems. Manually operated tail-first aircraft do not feel “sweet,” as the control forces are back to front. Ailerons tend to be heaviest, elevators lighter, and rudders (which are also weak in authority) are lightest of all.

All told, tail-first aircraft are only useful for cruising flight. This is the reason why the *Rutan Voyager*, which flew around the world without stopping in December 1986, had quite extreme tail-first features. Off-design (that is, when maneuvering, taking off, and landing) and without artificial aids, they are less useful overall than conventionally configured airplanes. The practical airplane started with the horizontal stabilizer in front, and it is worth asking why it migrated aft subsequently if it was not to some significant advantage.

To avoid the disadvantages of the tail-first configuration, some manufacturers are investigating three-surface arrangements (main wing in the middle, a foreplane, and a small but authoritative tail at the back) to bestow flexibility in handling qualities. We have yet to see if such an arrangement offers more cost-effective benefits than a conventional airplane built for the same task.

D. Structures and Materials

Traditionally, light airplanes have been constructed out of wooden frames covered with fabric and/or plywood, metal frames of welded steel tube covered in the same way as

those of wood, or metal frames with metal skins made out of aluminium alloys; these tend to be limited to single curvature panels.

Each of these is being replaced by composite structures of manmade materials, which decrease time spent in fabrication while enabling the constructor to make any desired shape.

The most convenient and durable way of constructing an amateur-built airframe is to form the primary, load-bearing structure of wood (Fig. 11). The aerodynamic shape is then given by filling in the open spaces between with rigid foam and covering all with either a fabric called Dynel or glass-reinforced plastic (GRP).

The use of load-bearing composites in structures has brought benefits and caused problems. Fibers are made variously of glass, aramid (Kevlar 49), or carbon set in matrices of resin. There are several significant differences between composite and metallic structures. The com-

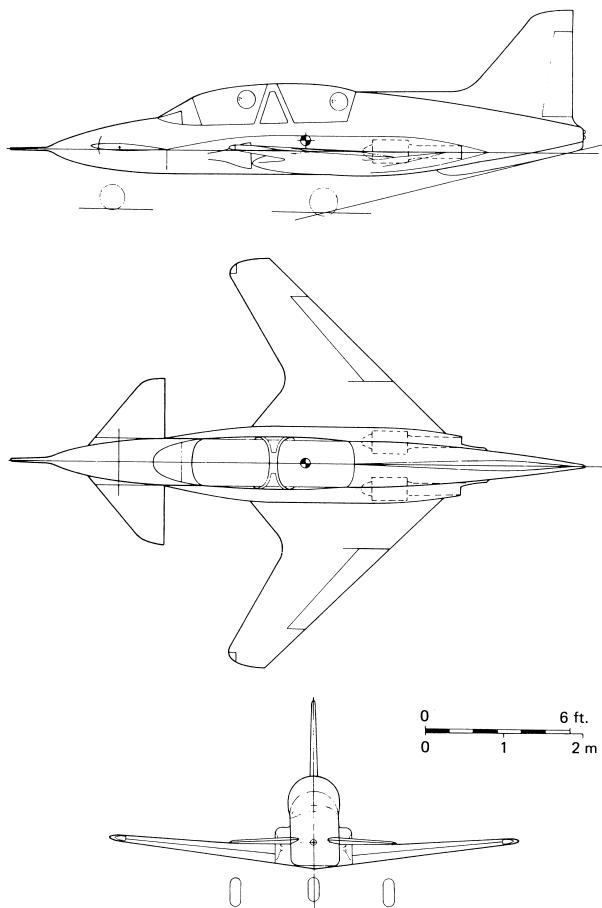


FIGURE 11 Designs for amateur-built, twin-jet light airplanes constructed of manmade composites to achieve M 0.65 (land plane) and 0.55 (seaplane). The land plane is a tail-first canard, while the seaplane has three sets of surfaces for improved handling qualities and a blended hull. [From Stinton, D. (1983), "The Design of the Aeroplane," Van Nostrand-Reinhold, Princeton, NJ.]

posite material is made at the same time as the component achieves its final form. Quality control of the process is therefore important. The formed laminate of fiber and resin has mechanical properties, which can be markedly different in different directions. They are said to be anisotropic.

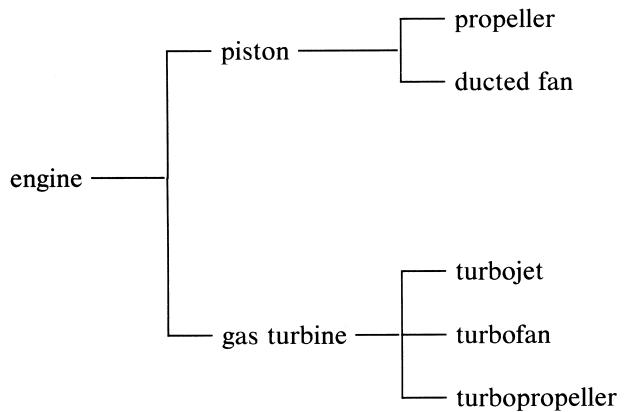
Composites can take less strain than metals before they fail and do not exhibit yielding and plastic behavior. Their somewhat fickle performance forces the stressman, out of prudence, to use additional factors of 1.2–1.8 on top of the usual minimum ultimate factor of safety of 1.5 in order to be safe. Environmental effects can have a marked influence on strength. A 1% absorption of water at 100°C reduces the strength in compression of a carbon fiber laminate by ~30%. Humidity, solar radiation, temperature, and cyclic loading can reduce strength significantly and not always in the expected direction.

While composites in general have better fatigue properties than, say, aluminium alloys, we do not know precisely how damage progresses through a structure, nor how to carry out an effective inspection. Impacts that cause barely visible cracking can cause fractures within the fibers and cracking of the resin some distance away. The strength of carbon fiber can be reduced by 50% by apparently light damage.

Composites are strong but tend to be brittle. They suffer erosion by rain and sand or dust to a greater extent than metals which allows moisture to get in. Allowance must be made for lower standards of repair control in the field.

E. Propulsion

Engines, which are the heart of any propulsion system, are devices for converting thermal energy, released by burning fuel, into propulsive work. General aviation engines can be grouped as follows:



The process of combustion is chemical, resulting in the emission of toxic exhaust products. Carbon monoxide and other hydrocarbons in the exhaust are reduced by

more efficient, hotter combustion. Leaning-off a piston engine (i.e., reducing the quantity of fuel mixed with air), within limits, can reduce fuel consumption by $\sim 20\%$ and raise working temperatures, which in turn reduces toxic emissions.

All propulsion systems are noisy because they are designed to sweep up a mass of more or less still air and accelerate it rearward at a high rate so as to produce forward thrust. Generally speaking, the larger the cross-sectional area of the working mass of air, the less acceleration needed to produce a given thrust. It is the final velocity, or true airspeed, of the working air, tearing and shearing its way through the surrounding quiet mass that causes most of the noise and offends the ear.

Thus, for a given thrust, a narrow jet of hot air squirted rearward by a jet engine is far noisier than a much slower wash of cold air worked on by a propeller. A fan engine is a compromise solution between the two: a hot gas generator, which is the heart of a jet engine, driving a larger diameter, quieter, and cooler running fan within a surrounding duct. Having said this, the conventional propeller, larger in diameter than a jet, is not necessarily quiet. The tip speeds of propeller blades must either be limited to no more than $\sim 60\%$ that of the local speed of sound (i.e., 0.6) or they must be made thin, so as to avoid or to reduce compressibility effects: the noisy shock waves caused by the overstressed air trying to readjust itself to its previously undisturbed ambient condition.

Propeller research is now part of the NASA GA program. Metal propellers have not changed significantly since the mid-1950s. Yet it is possible that noise levels, weight, and fatigue life might be improved by careful aerodynamic design of blade sections, reduction of tip speed, and the use of composite materials in their construction.

NASA's Lewis Research Center is sponsoring research on high-speed propellers for energy-efficient transport aircraft, designed to fly at about M 0.8. Inevitably the program will bring results that can be applied in GA—in particular, propellers for turboprop engines.

Pusher propellers can be more efficient than tractors. The ideal way of dumping engine exhaust to the atmosphere is by means of an annular exit around the propeller, but this is difficult to achieve in practice.

The propeller is one of the noisiest items. If the plane of rotation is in line with an occupant, it causes the occupant misery, especially when propeller tip speed and Mach number are high. Rear-mounted engines and propellers are welcome in these respects, because GA cabins are noisier on the whole than those of airliners. But a pusher propeller can be more offensively noisy for the world at large because of blades thumping unevenly through the wake of the airframe ahead. Wake velocity and pressure vary considerably around the path and radius of a blade.

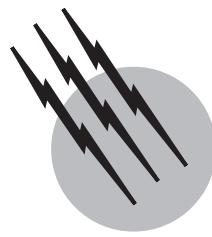
Body-mounted engines are worse sources of vibration than wing-mounted ones. Vibrations are transmitted straight into the structure surrounding the occupants. In general, the longer the *structural* path, which is not necessarily the shortest distance, between an occupant and an engine, the better the damping and reduction of vibration.

SEE ALSO THE FOLLOWING ARTICLES

AIRCRAFT AERODYNAMIC BOUNDARY LAYERS • AIRCRAFT AVIONICS • AIRCRAFT PERFORMANCE AND DESIGN • AIRCRAFT SPEED AND ALTITUDE • FLIGHT (AERODYNAMICS)

BIBLIOGRAPHY

- Federal Aviation Administration. "Federal Aviation Regulations," FAA, Washington, D.C.
- Federal Aviation Administration (1982). "FAA Statistical Handbook of Aviation," FAA, Washington, D.C.
- Grangier, M. (1980). "Research programmes in general aviation." *Interavia* 2, 1980.
- International Civil Aviation Organization. "Digests of Statistics," International Civil Aviation Organizations.
- Joint Aviation Authorities, "JAR-23 Normal, Utility, Aerobatic and Commuter Category Aeroplanes," Civil Aviation Authority, London (also Washington, D.C.).
- Stickle, J. W. (1977). "Technical Highlights in General Aviation," AIAA 13th Annu. Meet., 1977. Am. Inst. Aeronaut. Astronaut., Washington, D.C.
- Stinton, D. (1983). "The Design of the Aeroplane," Granada Publishing Ltd., London/Van Nostrand-Reinhold, Princeton, NJ.
- Stinton, D. (1984). "What do we think of the new look business aircraft?" *Interavia* 7, 1984.



Flight (Aerodynamics)

John D. Anderson, Jr.

National Air and Space Museum, Smithsonian Institution

- I. Historical Perspective
- II. Basic Applications
- III. General Governing Equations
- IV. How Aerodynamics is Subdivided
- V. Inviscid Incompressible Flow
- VI. Inviscid Compressible Flow
- VII. Viscous Flow
- VIII. Two Modern Aerodynamic Inventions
- IX. Epilogue

GLOSSARY

- Drag** Component of aerodynamic force in the freestream direction.
- Euler equations** Governing equations of fluid flow, not including viscous effects.
- Hypersonic flow** Flow at high Mach numbers, generally greater than 5.
- Inviscid flow** Flow wherein viscous effects are negligible.
- Lift** Component of aerodynamic force perpendicular to the freestream direction.
- Navier–Stokes equations** Governing equations of fluid flow including all the viscous effects.
- Shock waves** Very thin compression regions across which the flow properties change almost discontinuously.
- Subsonic flow** Flow where the Mach number is everywhere less than 1.
- Supersonic flow** Flow where the Mach number is everywhere greater than 1.

Transition Phenomena associated with the change from laminar to turbulent flow.

Transonic flow Flow involving mixed regions of subsonic and supersonic flow.

Viscous flow Flow where the transport phenomena of mass diffusion, viscosity, and thermal conductivity are important.

AERODYNAMICS is the study of the motion of a fluid, usually air, and its associated interactions with solid surfaces in the flow. These surfaces may be aerodynamic bodies such as airplanes and missiles or the inside walls of ducts such as inside rocket nozzles and wind tunnels. In this article, some of the basic nomenclature and principles from aerodynamics are presented and discussed. The content emphasizes aerodynamic theory, but also examines some experimental aspects. The purpose is to present an overall “bird’s-eye” view of aerodynamics in a self-contained manner.

I. HISTORICAL PERSPECTIVE

After centuries of false starts and misguided ideas, the concept of the modern airplane was finally conceived by the Englishman Sir George Cayley (1773–1857) in 1799. Cayley proposed a *fixed* wing for generating lift, another *separate* mechanism for propulsion, and a combined horizontal and vertical tail for stability. This was in contrast to the prevailing concepts of machines or humans with flapping wings to emulate birds. Cayley's ideas were adopted and nurtured by scores of would-be aviators during the 19th century; however, all attempts to design a powered, manned, heavier-than-air aircraft were unsuccessful until December 17, 1903. On that date, Orville (1871–1948) and Wilbur (1867–1912) Wright achieved the first successful powered flight of a manned airplane on the sand dunes of Kill Devil Hills, 4 miles south of Kitty Hawk, NC. The first flight lasted for only 12 sec, and it covered a distance over the ground of 120 ft—far less than the length of a football field—but it initiated a period of intense aeronautical development carrying through the biplanes of World War I, the high-performance propeller-driven airplanes of World War II, the advent of jet propulsion in the 1940s, and the spectacular development of rockets and spacecraft that now marks the last quarter of the 20th century.

The parallel development of aerodynamics, the science on which successful flight was achieved, began in a serious fashion with the classical mechanics of Isaac Newton (1642–1727) as described in his famous “Principia” of 1687. However, Newton considered a fluid flow as a uniform, rectilinear stream of particles, which, upon striking a surface, would transfer their normal momentum to the surface but would preserve their tangential momentum. This led to Newton's famous sine-squared law, which stated that the aerodynamic force on an inclined surface varies as $\sin^2 \theta$, where θ is the angle between the incoming fluid stream and the surface. Newton's fluid model and the sine-squared law were not accurate for most fluid flow problems (although it is a reasonable approximation for some of the very high speed, hypersonic flows encountered in modern aerodynamics), and the science of aerodynamics had to wait for the pioneering theoretical insights of Daniel Bernoulli (1700–1783) and Leonard Euler (1707–1783). Beginning with Bernoulli and Euler, and expanded by the work of Louis M. Navier (1785–1836) in France and Sir George G. Stokes (1819–1903) in England, the fundamental mathematical equations of fluid flow on which aerodynamics is built were well understood by the end of the 19th century. The 20th century has witnessed a virtual explosion of aerodynamic theory and experiments, carrying from the low-speed flows associated with the Wright Brothers' airplane through the hypersonic flows associ-

ated with the modern space shuttle. This article will, in part, survey such aerodynamic principles and theory.

II. BASIC APPLICATIONS

The practical objectives of aerodynamics are (1) the prediction of forces and moments on, and heat transfer to, bodies moving through a fluid (usually air) and (2) the determination of flows moving internally through ducts, such as flow through wind tunnels and jet engines. Applications in item (1) come under the heading of *external aerodynamics* and in item (2) under the heading of *internal aerodynamics*.

The definitions of aerodynamic forces and moments are illustrated in Fig. 1. The way that nature imposes a force on a body moving through a stationary fluid (or alternatively a stationary body immersed in a moving fluid) is by means of the *pressure distribution* exerted by the fluid and acting in a direction locally perpendicular to the surface and the *shear stress distribution* created by the frictional nature of the fluid and acting in a direction locally tangential to the surface. The pressure and shear stress distributions, integrated over the entire surface of the body, yield a net resultant aerodynamic force R and moment M on the body, as shown in Fig. 1. The line connecting the leading and trailing edges of the body is the *chord line*, and the distance from the leading to the trailing edge measured

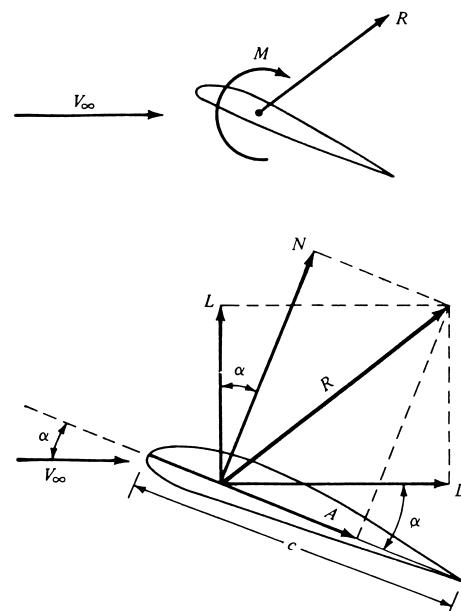


FIGURE 1 Aerodynamic forces and moments, illustrating the resolution of the resultant force into normal and axial forces and also into lift and drag. [From Anderson, J. D., Jr. (2001). “Fundamentals of Aerodynamics,” 3rd ed., McGraw-Hill, New York.]

along the chord line is simply called the *chord c*. The direction and magnitude of the flow far ahead of the body is the relative wind, or simply the freestream velocity V_∞ . The angle between c and V_∞ is the *angle of attack* α . The resultant aerodynamic force R is frequently resolved into two components, the *normal force* N and the *axial force* A , acting perpendicular and parallel to the chord, respectively. Even more frequently, R is resolved into the aerodynamic *lift* L and *drag* D , acting perpendicular and parallel to V_∞ , respectively. The moment M can be taken about any point, but is commonly taken about the leading edge (M_{LE}) or the quarter chord (a point $0.25c$ from the leading edge) ($M_{C/4}$).

The science of aerodynamics is more fundamentally based on *dimensionless parameters* involving force and moment coefficients rather than the forces and moments themselves. If ρ_∞ represents the freestream fluid density, such dimensionless parameters are

$$\text{Lift coefficient} \equiv C_L \equiv L/q_\infty S$$

$$\text{Drag coefficient} \equiv C_D \equiv D/q_\infty S$$

$$\text{Moment coefficient} \equiv C_M \equiv M/q_\infty Sc$$

where q_∞ is the freestream *dynamic pressure*, defined as $\frac{1}{2}\rho_\infty V_\infty^2$, and S is a suitable reference area of the body. These force and moment coefficients are in turn functions of (1) the shape of the body, (2) the angle of attack, and (3) various dimensionless similarity parameters. The most important similarity parameters are

$$\text{Mach number} = M_\infty/a_\infty$$

$$\text{Reynolds number} = Re_\infty = \rho_\infty V_\infty c/\mu_\infty$$

$$\text{Prandtl number} = Pr_\infty = \mu_\infty c_p/k_\infty$$

where a_∞ is the speed of sound in the free stream, μ_∞ is the freestream viscosity coefficient, c_p is the specific heat at constant pressure of the free stream, and k_∞ is the freestream thermal conductivity. For a body of given shape and angle of attack, C_L , C_D , and C_M are different functions of M_∞ , Re_∞ , and Pr_∞ .

The importance of these dimensionless parameters is seen in the principle of *flow similarity*. By definition, two different flows are dynamically similar if (1) the streamline patterns are geometrically similar; (2) the distributions of V/V_∞ , p/p_∞ , T/T_∞ , etc. throughout the flow field are the same when plotted against common nondimensional spatial coordinates (where V , p , and T are the local values of velocity, pressure, and temperature throughout the flow); and (3) C_L , C_D , and C_M are the same. Actually, item (3) is a consequence of item (2): if the nondimensional pressure and shear stress distributions over different bodies are the same, then the nondimensional force coefficients will be the same. The criteria that ensures that two different flows

are dynamically similar are (1) the bodies and any other solid boundaries are geometrically similar for both flows and (2) the similarity parameters (such as M_∞ , Re_∞ , and Pr_∞) are the same for both flows. The principle of flow similarity is used, for example, to compare results obtained in different wind tunnels at different speeds using different gases for different sized models: if the models are geometrically similar and at the same α , and if the similarity parameters such as M_∞ , Re_∞ , and Pr_∞ are the same for the two wind tunnels, then the measurements of C_L , C_D , and C_M from the two different wind tunnels will be the same. The same comparison can be made between wind-tunnel results for a model of a flight vehicle and the full-scale results for the actual flight vehicle moving through the atmosphere. If the free-flight values of M_∞ , Re_∞ , and Pr_∞ can be simultaneously simulated in the wind tunnel, then the measured values of C_L , C_D , and C_M in the wind tunnel will be the same as for full-scale flight in the atmosphere. This is the underlying justification for wind-tunnel testing in aerodynamics. Unfortunately, many wind tunnels are not able to simultaneously match the full-scale flight values of M_∞ , Re_∞ , and Pr_∞ , due to the many conflicting design variables necessary to achieve such simultaneous simulation. Instead, one wind tunnel may be able to simulate just M_∞ for free flight; another wind tunnel may be able to simulate just Re_∞ for full-scale flight. This is one of the reasons for the vast number of different types of wind tunnels that exists throughout the world today.

With the advent of very high speed flight, beginning in the 1950s with intercontinental ballistic missiles and continuing with the manned space program including the space shuttle and other hypersonic vehicles, the aerodynamic heating of the vehicle surface became a dominant aspect. Due to the vast amount of energy dissipated by friction in such flows, and the high temperatures encountered behind strong shock waves on such vehicles, heat-transfer rates to the vehicle surface can be very high. Aerodynamic heating, like aerodynamic forces and moments, is also governed by dimensionless parameters. If Q is the rate of heat added to the vehicle, then a dimensionless heat-transfer coefficient, the Stanton number, can be defined as

$$\text{Stanton number} \equiv C_H = Q/\rho_\infty V_\infty C_p(T_{aw} - T_w),$$

where T_w is the wall temperature of the surface and T_{aw} is the adiabatic wall temperature—temperature the wall would have if the heating rate were zero. In turn, $C_H = f(M_\infty, Re_\infty, Pr_\infty, T_w/T_\infty)$, i.e., the heat-transfer coefficient is a function of the similarity parameters, Mach number, Reynolds number, Prandtl number, and the wall-to-freestream temperature ratio. Hence, the principle of dynamic similarity also holds for aerodynamic heating.

III. GENERAL GOVERNING EQUATIONS

The governing equations of aerodynamics are based on three basic physical principles:

1. Mass is conserved.
2. Newton's second law:
force = (mass) × (acceleration).
3. Energy is conserved.

These physical principles are applied to various models of a fluid flow, for example, an infinitesimally small fluid element moving along a streamline. This leads to a series of partial differential equations relating the flowfield variables of density ρ ; pressure p ; three components of velocity in the x , y , and z directions in Cartesian space u , v , and w , respectively; and internal energy e (where e depends in part on temperature, e.g., for a calorically perfect gas $e = c_v T$ where c_v is the specific heat at constant volume). These equations are

Continuity

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{V}) = 0 \quad (1)$$

Momentum (x component)

$$\rho \frac{Du}{Dt} = -\frac{\partial p}{\partial x} + \frac{\partial \tau_{xx}}{\partial x} + \frac{\partial \tau_{yx}}{\partial y} + \frac{\partial \tau_{zx}}{\partial z} \quad (2)$$

Momentum (y component)

$$\rho \frac{Dv}{Dt} = -\frac{\partial p}{\partial y} + \frac{\partial \tau_{xy}}{\partial x} + \frac{\partial \tau_{yy}}{\partial y} + \frac{\partial \tau_{zy}}{\partial z} \quad (3)$$

Momentum (z component)

$$\rho \frac{Dw}{Dt} = -\frac{\partial p}{\partial z} + \frac{\partial \tau_{xz}}{\partial x} + \frac{\partial \tau_{yz}}{\partial y} + \frac{\partial \tau_{zz}}{\partial z} \quad (4)$$

Energy

$$\begin{aligned} \rho \frac{D(e + V^2/2)}{Dt} &= p\dot{q} + \frac{\partial}{\partial x} \left(k \frac{\partial T}{\partial x} \right) + \frac{\partial}{\partial y} \left(k \frac{\partial T}{\partial y} \right) \\ &\quad + \frac{\partial}{\partial z} \left(k \frac{\partial T}{\partial z} \right) - \nabla \cdot p \mathbf{V} + \frac{\partial(u\tau_{xx})}{\partial x} \\ &\quad + \frac{\partial(u\tau_{yx})}{\partial y} + \frac{\partial(u\tau_{zx})}{\partial z} + \frac{\partial(v\tau_{xy})}{\partial x} \\ &\quad + \frac{\partial(v\tau_{yy})}{\partial y} + \frac{\partial(v\tau_{zy})}{\partial z} + \frac{\partial(w\tau_{xz})}{\partial x} \\ &\quad + \frac{\partial(w\tau_{yz})}{\partial y} + \frac{\partial(w\tau_{zz})}{\partial z} \end{aligned} \quad (5)$$

where

$$\tau_{xy} = \tau_{yx} = \mu \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right)$$

$$\tau_{yz} = \tau_{zy} = \mu \left(\frac{\partial w}{\partial y} + \frac{\partial v}{\partial z} \right)$$

$$\tau_{zx} = \tau_{xz} = \mu \left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x} \right)$$

$$\tau_{xx} = \lambda(\nabla \cdot \mathbf{V}) + 2\mu \frac{\partial u}{\partial x}$$

$$\tau_{yy} = \lambda(\nabla \cdot \mathbf{V}) + 2\mu \frac{\partial v}{\partial y}$$

$$\tau_{zz} = \lambda(\nabla \cdot \mathbf{V}) + 2\mu \frac{\partial w}{\partial z}$$

\dot{q} = rate of volumetric heat added per unit mass of gas (say by radiation)

k = thermal conductivity

$$\mathbf{V} = u\mathbf{i} + v\mathbf{j} + w\mathbf{k}$$

$$\begin{aligned} \frac{D}{Dt} &= \text{substantial derivative operator} \\ &= \frac{\partial}{\partial t} + u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y} + w \frac{\partial}{\partial z} \end{aligned}$$

These equations describe the general unsteady, three-dimensional flow of a compressible viscous fluid. They are the foundations of theoretical aerodynamics. Historically, Eqs. (2)–(4) are called the *Navier–Stokes equations*; however, in recent aerodynamic literature, especially dealing with computational solutions, the entire system of Eqs. (1)–(5) has been called the Navier–Stokes equations.

Equations (1)–(5) are nonlinear, coupled, partial differential equations with no known analytical solution. Therefore, the historical development of theoretical aerodynamics has utilized various simpler models of flows, which are governed approximately by special, simplified forms of Eqs. (1)–(5). For this reason and others, aerodynamics is conveniently subdivided into various types of flows, as described in the next section.

IV. HOW AERODYNAMICS IS SUBDIVIDED

An understanding of aerodynamics, like that of any other physical science, is obtained through a “building-block” approach—we dissect the discipline, form the parts into nice polished blocks of knowledge, and then later attempt to reassemble the blocks to form an understanding of the whole. An example of this process is the way that different types of aerodynamic flows are categorized and visualized. Although nature has no trouble setting up the most detailed

and complex flow with a whole spectrum of interacting physical phenomena, we must attempt to understand such flows by modeling them with less detail and neglecting some of the (hopefully) less significant phenomena. As a result, a study of aerodynamics has evolved into a study of numerous and distinct types of flow. The purpose of this section is to itemize and contrast these types of flow and to briefly describe their most important physical phenomena.

A. Continuum versus Free-Molecule Flow

Consider the flow over a body, say, for example, a circular cylinder of diameter d . Also, consider the fluid to consist of individual molecules, which are moving about in random motion. The mean distance that a molecule travels between collisions with neighboring molecules is defined as the mean free path λ . If λ is orders of magnitude smaller than the scale of the body measured by d , then the flow appears to the body as a continuous substance. The molecules impact the body surface so frequently that the body cannot distinguish the individual molecular collisions, and the surface feels the fluid as a continuous medium. Such flow is called continuum flow. The other extreme is where λ is on the same order as the body scale; here, the gas molecules are spaced so far apart (relative to d) that collisions with the body surface occur only infrequently, and the body surface can feel distinctly each molecular impact. Such flow is called free molecular flow. For manned flight, vehicles such as the space shuttle encounter free molecular flow at the extreme outer edge of the atmosphere, where the air density is so low that λ becomes on the order of the shuttle size. There are intermediate cases, where flows can exhibit some characteristics of both continuum and free-molecule flows; such flows are generally labeled “low-density flows,” in contrast to continuum flow. By far, the vast majority of practical aerodynamic applications involves continuum flows. Low-density and free-molecule flows are just a small part of the total spectrum of aerodynamics.

B. Inviscid versus Viscous Flow

A major facet of a gas or liquid is the ability of the molecules to move rather freely. When the molecules move, even in a very random fashion, they obviously transport their mass, momentum, and energy from one location to another in the fluid. This transport on a molecular scale gives rise to the phenomena of mass diffusion, viscosity (friction), and thermal conduction. All real flows exhibit the effects of these transport phenomena; such flows are called viscous flows. In contrast, a flow that is assumed to involve no friction, thermal conduction, or diffusion is called an inviscid flow. Inviscid flows do not truly exist

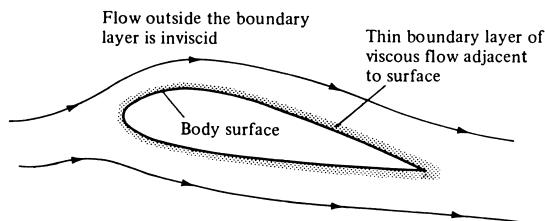


FIGURE 2 Division of an aerodynamic flow into a viscous boundary layer adjacent to the surface and an inviscid outer flow. [From Anderson, J. D., Jr. (2001). "Fundamentals of Aerodynamics," 3rd ed., McGraw-Hill, New York.]

in nature; however, there are many practical aerodynamic flows (more than you would think) where the influence of transport phenomena is small, and we can model the flow as being inviscid. For this reason, more than 70% of aerodynamic theory deals with inviscid flows.

Theoretically, inviscid flow is approached in the limit as the Reynolds number goes to infinity. However, for practical problems, many flows with high but finite Re can be assumed to be inviscid. For such flows, the influence of friction, thermal conduction, and diffusion is limited to a very thin region adjacent to the body surface (the boundary layer), and the remainder of the flow outside this thin region is essentially inviscid. This division of the flow into two regions is illustrated in Fig. 2. For flows over slender bodies, such as the airfoil sketched in Fig. 2, inviscid theory adequately predicts the pressure distribution and lift on the body and gives a valid representation of the streamlines and flow field away from the body. However, because friction (shear stress) is a major source of aerodynamic drag, inviscid theories by themselves cannot adequately predict total drag.

In contrast, there are some flows that are dominated by viscous effects. For example, if the airfoil in Fig. 2 is inclined to a high incidence angle to the flow (high angle of attack), then the boundary layer will tend to separate from the top surface, and a large wake is formed downstream. The separated flow is sketched at the top of Fig. 3; it is characteristic of the flow field over a “stalled” airfoil. Separated flow also dominates the aerodynamics of blunt bodies, such as the cylinder at the bottom of Fig. 3. Here, the flow expands around the front face of the cylinder, but separates from the surface on the rear face, forming a rather fat wake downstream. The types of flow illustrated in Fig. 3 are dominated by viscous effects: no inviscid theory can independently predict the aerodynamics of such flows.

C. Incompressible versus Compressible Flows

A flow in which the density ρ is constant is called incompressible. In contrast, a flow where the density is variable

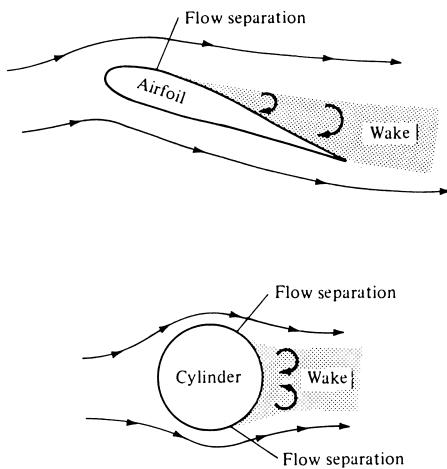


FIGURE 3 Illustration of flow separation. [From Anderson, J. D., Jr. (2001). “Fundamentals of Aerodynamics,” 3rd ed., McGraw-Hill, New York.]

is called compressible. For our purposes here, we will simply note that all flows, to a greater or lesser extent, are compressible: truly incompressible flow, where the density is precisely constant, does not occur in nature. However, analogous to our discussion of inviscid flow, there are a number of aerodynamic problems that can be modeled as being incompressible without any detrimental loss of accuracy. For example, the flow of homogenous liquids is treated as incompressible, and hence, most problems involving hydrodynamics assume ρ is a constant. Also, the flow of gases at low Mach number is essentially incompressible; for $M < 0.3$, it is always safe to assume ρ is a constant. This was the flight regime of all airplanes from the Wright Brothers’ first flight in 1903 to just prior to World War II. It is still the flight regime of most small, general aviation aircraft of today. Hence, there exists a large bulk of aerodynamic experimental and theoretical data for incompressible flows. On the other hand, high-speed flow (near Mach 1 and above) must be treated as compressible; for such flows, ρ can vary over wide latitudes.

D. Mach-Number Regimes

Of all the ways of subdividing and describing different aerodynamic flows, the distinction based on Mach number is probably the most prevalent. If M is the local Mach number at an arbitrary point in a flow field, then by definition the flow is locally

$$\text{Subsonic if } M < 1$$

$$\text{Sonic if } M = 1$$

$$\text{Supersonic if } M > 1$$

Looking at the whole flow field simultaneously, four different speed regimes can be identified using Mach number as the criterion.

1. Subsonic flow ($M < 1$ everywhere). A flow field is defined as subsonic if the Mach number is less than 1 at every point. Subsonic flows are characterized by smooth streamlines (no discontinuity in slope), as sketched in Fig. 4a. Moreover, since the flow velocity is everywhere less than the speed of sound, disturbances in the flow (say the sudden deflection of the trailing edge of the airfoil in Fig. 4a) propagate both upstream and downstream and are felt throughout the entire flow field. Note that a freestream Mach number M_∞ less than 1 does not guarantee a totally subsonic flow over the body. In expanding over an aerodynamic shape, the flow velocity increases above the freestream value, and if M_∞ is close enough to 1, the local Mach number may become supersonic in certain regions of the flow. This gives rise to a rule of thumb that $M_\infty < 0.8$ for subsonic flow over slender bodies. For blunt bodies, M_∞ must be even lower to insure totally subsonic flow. (Again, emphasis is made that the above is just a loose rule of thumb and should not be taken as a precise quantitative definition.) Also, note that incompressible

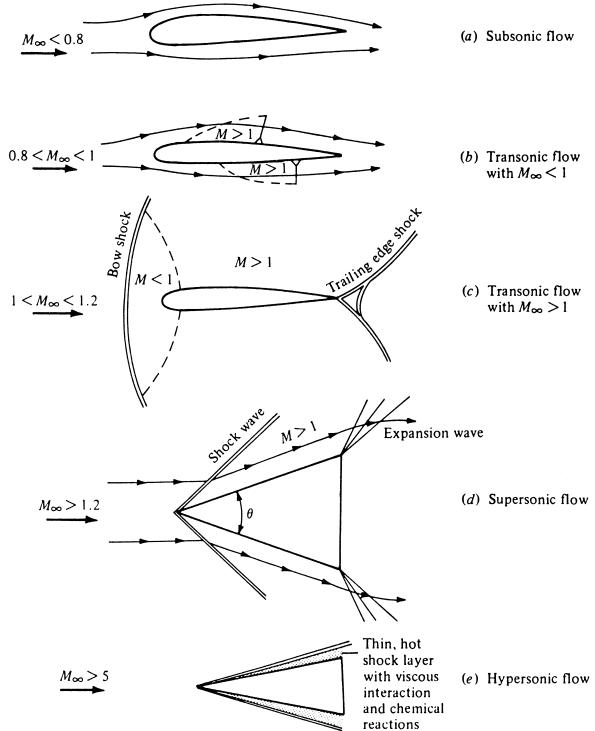


FIGURE 4 Characterization of flow fields based on the Mach number range. [From Anderson, J. D., Jr. (2001). “Fundamentals of Aerodynamics,” 3rd ed., McGraw-Hill, New York.]

flow is a special limiting case of subsonic flow where $M \rightarrow 0$.

2. Transonic flow (mixed regions where $M < 1$ and $M > 1$). As stated above, if M_∞ is subsonic but near unity, the flow can become locally supersonic ($M > 1$). This is sketched in Fig. 4b, which shows pockets of supersonic flow over both the top and the bottom surfaces of the airfoil, terminated by weak shock waves behind which the flow becomes subsonic again. Moreover, if M_∞ is increased slightly above unity, a bow shock wave is formed in front of the body; behind this shock wave the flow is locally subsonic, as shown in Fig. 4c. This subsonic flow subsequently expands to a low supersonic value over the airfoil. Weak shock waves are usually generated at the trailing edge, sometimes in a “fishtail” pattern as shown in Fig. 4c. The flow fields shown in Figs. 4b and 4c are characterized by mixed subsonic–supersonic flows and are dominated by the physics of both types of flows. Hence, such flow fields are called transonic flows. Again, as a rule of thumb for slender bodies, transonic flows occur for freestream Mach numbers in the range $0.8 < M_\infty < 1.2$.

3. Supersonic flow ($M > 1$ everywhere). A flow field is defined as supersonic if the Mach number is greater than 1 at every point. Supersonic flows are frequently characterized by the presence of shock waves across which the flow properties and streamlines change discontinuously (in contrast to the smooth, continuous variations in subsonic flows). This is illustrated in Fig. 4d for supersonic flow over a sharp-nosed wedge; the flow remains supersonic behind the oblique shock wave from the tip. Also shown are distinct expansion waves, which are common in supersonic flow. (Again, the listing of $M_\infty > 1.2$ is strictly a rule of thumb. For example, in Fig. 4d, if θ is made large enough, the oblique shock wave will detach from the tip of the wedge and will form a strong, curved bow shock ahead of the wedge with a substantial region of subsonic flow behind the wave. Hence, the totally supersonic flow sketched in Fig. 4d is destroyed if θ is too large for a given M_∞ . This shock detachment phenomenon can occur at any value of $M_\infty > 1$, but the value of θ at which it occurs increases as M_∞ increases. In turn, if θ is made infinitesimally small, the flow field in Fig. 4d holds for $M_\infty \geq 1.0$. However, the above discussion clearly shows that the listing of $M_\infty > 1.2$ in Fig. 4d is a very tenuous rule of thumb and should not be taken literally. In a supersonic flow, because the local flow velocity is greater than the speed of sound, disturbances created at some point in the flow cannot work their way upstream (in contrast to subsonic flow). This property is one of the most significant physical differences between subsonic and supersonic flows. It is the basic reason why shock waves occur in supersonic flows but do not occur in steady subsonic flow.

4. Hypersonic flow (very high supersonic speeds). Refer again to the wedge in Fig. 4d. Assume θ is a given, fixed value. As M_∞ increases above 1, the shock wave moves closer to the body surface. Also, the strength of the shock wave increases, leading to higher temperatures in the region between the shock and the body (the shock layer). If M_∞ is sufficiently large, the shock layer becomes very thin, and interactions between the shock wave and the viscous boundary layer on the surface occur. Also, the shock layer temperature becomes high enough that chemical reactions occur in the air. The O_2 and N_2 molecules are torn apart; that is, the gas molecules dissociate. When M_∞ becomes large enough such that viscous interaction and/or chemically reacting effects begin to dominate the flow (Fig. 4e), the flow field is called hypersonic. (Again, a somewhat arbitrary but frequently used rule of thumb for hypersonic flow is $M_\infty > 5$.) Hypersonic aerodynamics received a great deal of attention during the period 1955–1970 because atmospheric entry vehicles encounter the atmosphere at Mach numbers between 25 (ICBMs) and 36 (the Apollo lunar return vehicle.) From 1985 to the present, this attention has shifted to air-breathing hypersonic cruise vehicles and single-stage-to-orbit vehicles. Today, hypersonic aerodynamics is just part of the whole spectrum of realistic flight speeds.

In summary, we attempt to organize our study of aerodynamic flows according to one or more of the various categories discussed in this section. The block diagram in Fig. 5 is presented to help emphasize these categories and to show how they are related.

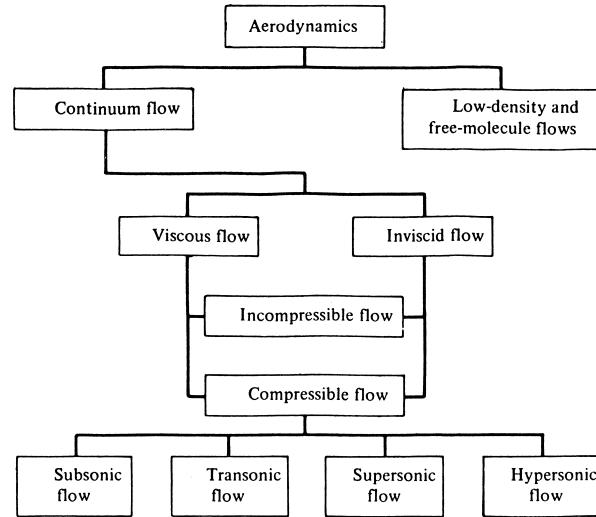


FIGURE 5 Block diagram for the various categories of aerodynamic flows. [From Anderson, J. D., Jr. (2001). “Fundamentals of Aerodynamics,” 3rd ed., McGraw-Hill, New York.]

There are several additional types of flows frequently assumed in aerodynamics.

1. Irrotational flow. Flow where the individual elements of fluid moving along a streamline are in translational motion only and are not rotating. For irrotational flow, a scalar potential function $\phi = \phi(x, y, z)$ can always be defined in such a fashion that

$$\mathbf{V} = \nabla\phi.$$

This is an important aspect of theoretical aerodynamics; ϕ is called the *velocity potential*.

2. Adiabatic flow. Flow where no heat is added or taken away.

3. Isentropic flow. Flow undergoing a thermodynamic process that is both adiabatic and reversible. For such flow, the thermodynamic entropy is constant along a streamline.

V. INVISCID INCOMPRESSIBLE FLOW

Incompressible flow is constant-density flow. Inviscid flow is flow where the transport phenomena of mass diffusion, viscosity, and thermal conduction are negligibly small. Under these two assumptions, the governing equations of motion, Eqs. (1–5), become considerably simpler. Equation (1) reduces to

$$\nabla \cdot \mathbf{V} = 0, \quad (6)$$

which is the continuity equation for incompressible flow. Eqs. (2)–(4) combine and reduce to

$$p_1 + \frac{1}{2}\rho V_1^2 = p_2 + \frac{1}{2}\rho V_2^2. \quad (7)$$

Equation (7) is the famous Bernoulli equation, and it relates the pressure and velocity at two different points along a streamline. Another way of writing the Bernoulli equation is

$$p + \frac{1}{2}\rho V^2 = \text{constant} \quad (8)$$

along a streamline. If all the streamlines of the flow originate from a uniform reservoir (such as the reservoir of a wind tunnel or the atmosphere far ahead of a moving flight vehicle), then the constant in Eq. (8) is the same for all streamlines, and Eq. (7) then relates the pressure and velocity at *any* two points in the flow, not necessarily on the same streamline. For inviscid incompressible flow, the energy equation [Eq. (5)] will also reduce to the Bernoulli equation; hence, the energy equation for such a flow is redundant and is not needed. Therefore, the Bernoulli equation can be viewed as both a statement of Newton's second law, $F = ma$, and a statement of the conservation of mechanical energy in an inviscid incompressible flow.

The Bernoulli equation is used in many applications of low-speed aerodynamics: the flow over airfoils, the flow through ducts, and the utilization of the Pitot tube for airspeed measurements, to name just a few. Let us examine more closely the last item mentioned above, namely, the Pitot tube.

In 1732, the Frenchman Henri Pitot was busy trying to measure the flow velocity of the Seine River in Paris. One of the instruments he used was his own invention—a strange-looking tube bent into an \angle shape, as shown in Fig. 6. Pitot oriented one of the open ends of the tube so that it faced directly into the flow. In turn, he used the pressure inside this tube to measure the water flow velocity. This was the first time in history that a proper measurement of fluid velocity was made, and Pitot's invention has carried through to the present day as the Pitot tube—one of the most common and frequently used instruments in any modern aerodynamic laboratory. Moreover, a Pitot tube is the most common device for measuring flight velocities of airplanes. The purpose of this section is to describe the basic principle of the Pitot tube.

Consider a flow with pressure p_1 moving with velocity V_1 , as sketched at the left of Fig. 6. Let us consider the significance of the pressure p_1 more closely. The pressure is associated with the time rate of change of momentum of the gas molecules impacting on or crossing a surface; that is, pressure is clearly related to the motion of the molecules. This motion is very random, with molecules moving in all direction with various velocities. Now imagine that you hop on a fluid element of the flow and ride with it at the velocity V_1 . The gas molecules, because of their random motion, will still bump into you, and you will

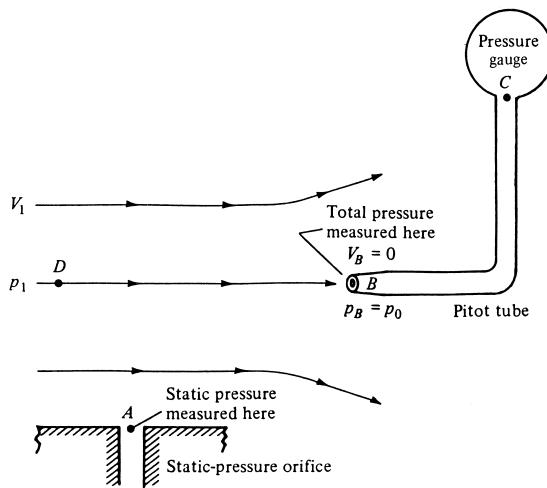


FIGURE 6 Illustration of Pitot and static pressure measurements. [From Anderson, J. D., Jr. (2001). "Fundamentals of Aerodynamics," 3rd ed., McGraw-Hill, New York.]

feel the pressure p_1 of the gas. We now give this pressure a specific name: the static pressure. Static pressure is a measure of the purely random motion of molecules in the gas; it is the pressure you feel when you ride along with the gas at the local flow velocity. All pressures used in this article so far have been static pressures; the pressure p appearing in all our previous equations has been the static pressure. In engineering, whenever a reference is made to “pressure” without further qualification, that pressure is always interpreted as the static pressure. Furthermore, consider a boundary of the flow, such as a wall, where a small hole is drilled perpendicular to the surface. The plane of the hole is parallel to the flow, as shown at point A in [Fig. 6](#). Because the flow moves over the opening, the pressure felt at point A is due only to the random motion of the molecules; that is, at point A, the static pressure is measured. Such a small hole in the surface is called a static pressure orifice, or a static pressure tap.

In contrast, consider that a Pitot tube is now inserted into the flow, with an open end facing directly into the flow. That is, the plane of the opening of the tube is perpendicular to the flow, as shown at point B in [Fig. 6](#). The other end of the Pitot tube is connected to a pressure gauge, such as point C in [Fig. 6](#); that is, the Pitot tube is closed at point C. For the first few milliseconds after the Pitot tube is inserted into the flow, the gas will rush into the open end and will fill the tube. However, the tube is closed at point C; there is no place for the gas to go, and hence after a brief period of adjustment, the gas inside the tube will stagnate; that is, the gas velocity inside the tube will go to zero. Indeed, the gas will eventually pile up and stagnate everywhere inside the tube, including at the open mouth at point B. As a result, the streamline of the flow that impinges directly at the open face of the tube (streamline DB in [Fig. 6](#)) sees this face as an obstruction to the flow. The fluid elements along streamline DB slow down as they get closer to the Pitot tube and go to zero velocity right at point B. Any point in a flow where $V = 0$ is called a stagnation point of the flow; hence, point B at the open face of the Pitot tube is a stagnation point, where $V_B = 0$. In turn, from Bernoulli’s equation we know the pressure increases as the velocity decreases. Hence, $p_B > p_1$. The pressure at a stagnation point is called the stagnation pressure, or total pressure, denoted by p_0 . Hence, at point B, $p_B = p_0$.

From the above discussion, we see that two types of pressure can be defined for a given flow: static pressure, which is the pressure you feel by moving with the flow at its local velocity V_1 , and total pressure, which is the pressure that the flow achieves when the velocity is reduced isentropically to zero.

How is the Pitot tube used to measure flow velocity? To answer this question, first note that the total pressure p_0 exerted by the flow at the tube inlet (point B) is impressed

throughout the tube (there is no flow inside the tube; hence, the pressure everywhere inside the tube is p_0). Therefore, the pressure gage at point C reads p_0 . This measurement, in conjunction with a measurement of the static pressure p_1 at point A, yields the difference between total and static pressure, $p_0 - p_1$, and it is this pressure difference that allows the calculation of V_1 via Bernoulli’s equation. In particular, apply Bernoulli’s equation between point A, where the pressure and velocity are p_1 and V_1 , respectively, and point B, where the pressure and velocity are p_0 and $V = 0$, respectively:

$$p_A + \frac{1}{2}\rho V_A^2 = p_B + \frac{1}{2}\rho V_B^2$$

or

$$p_1 + \frac{1}{2}\rho V_1^2 = p_0 + 0.$$

Solving for V_1 , we have

$$V_1 = \sqrt{\frac{2(p_0 - p_1)}{\rho}}.$$

The above equation allows the calculation of velocity simply from the measured difference between total and static pressure. The total pressure p_0 is obtained from the Pitot tube, and the static pressure p_1 is obtained from a suitably placed static pressure tap.

It is possible to combine the measurement of both total and static pressure in one instrument, a Pitot-static probe. A Pitot-static probe measures p_0 at the nose of the probe and p_1 at a suitably placed static pressure tap on the probe surface downstream of the nose.

In Eq. (8), the grouping $\frac{1}{2}\rho V^2$ is called the *dynamic pressure* by definition and is used in all flows, incompressible to hypersonic:

$$\text{Dynamic pressure} \equiv \frac{1}{2}\rho V^2.$$

However, for incompressible flow, the dynamic pressure has special meaning; it is precisely the difference between total and static pressure:

$$q = p_0 - p.$$

It is important to keep in mind that this relation comes from Bernoulli’s equation and thus holds for incompressible flow only. For compressible flow, where Bernoulli’s equation is not valid, the pressure difference $p_0 - p$ is not equal to q . The velocities of compressible flows, both subsonic and supersonic, can be measured by means of a Pitot tube, but the equations are different from the above. (Velocity measurements in subsonic and supersonic compressible flows are discussed in Section VI.)

In aerodynamics, it is useful to define a dimensionless pressure as

$$C_p \equiv \frac{p - p_\infty}{q_\infty}, \quad (9)$$

where $q_\infty = \frac{1}{2} \rho_\infty V_\infty^2$ and where C_p is defined as the *pressure coefficient* used throughout aerodynamics, from incompressible to hypersonic flow. In the aerodynamic literature it is very common to find pressures given in terms of C_p rather than the pressure itself. For *incompressible flow*, C_p can be expressed in terms of velocity only. By applying the Bernoulli equation to Eq. (9), the result for incompressible flow is

$$C_p = 1 - (V/V_\infty)^2. \quad (10)$$

The pressure coefficient is more than just a dimensionless pressure; it is a similarity parameter in the same sense as C_L and C_D , and for a given body shape, angle of attack, and geometrically similar location in the flow, $C_p = f(M_\infty, Re_\infty, Pr_\infty)$.

In general treatments of inviscid, incompressible, irrotational flows, the basic governing equations reduce to Laplace's equation in terms of the velocity potential. Combining Eq. (6) with the irrotational expression $\mathbf{V} = \nabla\phi$, we directly obtain

$$\nabla^2\phi = \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2} = 0, \quad (11)$$

which is *Laplace's equation*—one of the most familiar equations from mathematical physics. Hence, solutions of incompressible, inviscid, irrotational flows are called *potential flow solutions*.

In theoretical aerodynamics, the concept of the stream function is sometimes employed. By definition, the stream function $\bar{\psi}$ is a constant along a given streamline, and the change in $\bar{\psi}$, $\Delta\bar{\psi}$, between two streamlines is equal to the *mass flow* between these streamlines. For incompressible flow, a modified stream function ψ is defined such that $\Delta\bar{\psi}$ between two streamlines equals the *volume flow* between the streamlines. From the definition of the stream function, it can be shown that the flow velocity can be found by differentiating ψ , for example,

$$\nabla^2\psi = \frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} = 0. \quad (12)$$

Hence, for such flows, Laplace's equation governs both ϕ and ψ . Moreover, in a flow field, lines of constant ψ (i.e., streamlines) are everywhere orthogonal to lines of constant ϕ (i.e., equipotential lines).

Laplace's equation is linear, and hence, any number of particular solutions, say ϕ_1, ϕ_2, ϕ_3 , and ϕ_4 , can be added to obtain another solution, say ϕ_5 :

$$\phi_5 = \phi_1 + \phi_2 + \phi_3 + \phi_4.$$

This allows the solution of a given potential flow to be synthesized by superimposing a number of other, more elementary, flows. This is the major underlying strategy of the solution of inviscid, incompressible, irrotational flows. In the synthesis of such flows, the following elementary flows are useful.

1. Uniform flow. Constant-property flow at velocity V_∞ , with straight streamlines oriented in a single direction (say, the x direction),

$$\begin{aligned}\phi_1 &= V_\infty x = V_\infty r \cos \theta \\ \psi_1 &= V_\infty y = V_\infty r \sin \theta\end{aligned}$$

where r and θ are polar coordinates.

2. Source flow (in two dimensions). Streamlines are straight lines emanating from a central point (say, the origin of the coordinate system) where the velocity along each streamline varies inversely with distance from the central point. In polar coordinates,

$$\phi_2 = \frac{\Lambda}{2\pi} \ln r \quad \psi_2 = \frac{\Lambda}{2\pi} \theta,$$

where Λ is the source strength defined as the rate of volume flow from the source. A negative value of Λ denotes a *sink*, where the flow moves toward the central point.

3. Source flow (in three dimensions). A flow with straight streamlines emanating in three dimensions from an origin, where the velocity varies inversely as the square of the distance from the origin, and

$$\phi_3 = -\frac{\lambda}{4\pi r},$$

where λ is the source strength, that is, the rate of volume flow from the origin. For a sink, λ is negative.

4. Doublet flow (in two dimensions). The superposition of a source and sink of equal but opposite strength, Λ and $-\Lambda$, where the distance l between the two approaches zero at the same time that the product Λl remains constant. The strength of the doublet is $\kappa \equiv \Lambda l$. In polar coordinates,

$$\phi_4 = \left(\frac{\kappa}{2\pi} \right) \frac{\cos \theta}{r} \quad \psi_4 = -\left(\frac{\kappa}{2\pi} \right) \frac{\sin \theta}{r}.$$

5. Doublet flow (in three dimensions). The superposition of a three-dimensional source and sink of equal but opposite strength, λ and $-\lambda$, where the distance l between the two approaches zero at the same time that the product $\mu = \lambda l$ remains constant. In spherical coordinates,

$$\phi_5 = -\left(\frac{\mu}{4\pi} \right) \frac{\cos \theta}{r^2}.$$

6. Vortex flow (in two dimensions). Streamlines are concentric circles about a given central point, where the velocity along any given circular streamline is constant

but varies from one streamline to another inversely with distance from the common center. In polar coordinates with an origin at the central point,

$$\phi_6 = -\frac{\Gamma}{2\pi} \theta \quad \psi_6 = \frac{\Gamma}{2\pi} \ln r,$$

where Γ is the vortex strength, defined as the circulation about the vortex.

The term “circulation” introduced above is a general concept in aerodynamics and is particularly useful in the analysis of low-speed airfoils and wings. Circulation is defined as

$$\Gamma = -\oint \mathbf{V} \cdot d\mathbf{s}.$$

That is, circulation is the line integral of flow velocity integrated about the closed curve C drawn in the flow. For vortex flow, the vortex strength Γ is the circulation taken about any closed curve that encloses the central point.

The six elementary flows described above, by themselves, are not practical flow fields. However, they can be superimposed in various ways to synthesize practical flows in two and three dimensions, such as flows over cylinders, spheres, airfoils, wings, and whole airplanes. For example, three such flows are described below. Keep in mind that the expressions given for ϕ and ψ for the elementary flows are solutions of Laplace’s equation for those flows, and therefore, such solutions can be added together (superimposed) to produce other solutions.

1. Nonlifting flow over a circular cylinder. This symmetrical flow is synthesized by the superposition of a uniform flow with a doublet, yielding the stream function for a cylinder of radius R in a free stream of velocity V_∞ :

$$\psi = (V_\infty r \sin \theta) \left(1 - \frac{R^2}{r^2} \right).$$

The resulting pressure-coefficient variation over the surface of the cylinder is symmetrical and is given by

$$C_p = 1 - 4 \sin^2 \theta.$$

Because the pressure variation is symmetrical, there is no lift or drag theoretically predicted for the cylinder.

2. Lifting flow over a circular cylinder. In actual experience, a circular cylinder spinning about its axis and immersed in a flow with velocity V_∞ will experience a lift force. This is because the frictional effect between the fluid and the spinning cylinder tends to increase the flow velocity on one side of the cylinder and to decrease the flow velocity on the other side. From the Bernoulli equation, these unequal velocities lead to unequal pressures on both sides of the cylinder, causing a lift force to be produced.

This phenomenon is sometimes called the “Magnus effect.” The curve of a spinning baseball and the hook or slice of a spinning golf ball are examples of the Magnus effect. In this case, for a cylinder of radius R in a flow with velocity V_∞ , the lifting flow is synthesized by the superposition of a uniform flow, a doublet, and a vortex, yielding

$$\psi = (V_\infty r \sin \theta) \left(1 - \frac{R^2}{r^2} \right) + \frac{\Gamma}{2\pi} \ln \frac{r}{R}.$$

The resulting pressure coefficient distribution is not symmetrical and is given by

$$C_p = 1 - \left[4 \sin^2 \theta + \frac{2\Gamma \sin \theta}{\pi R V_\infty} + \left(\frac{\Gamma}{2\pi R V_\infty} \right)^2 \right].$$

In turn, this yields a theoretical value for lift and drag per unit span of the cylinder as

$$L = \rho_\infty V_\infty \Gamma \quad D = 0,$$

where Γ is the value of the circulation about the cylinder and is also equal to the strength of the elementary vortex used in the superposition process.

3. Nonlifting flow over a sphere. This flow is synthesized by superposition of a uniform flow and a three-dimensional doublet, yielding a result for the pressure-coefficient variation over the spherical surface as

$$C_p = 1 - \frac{9}{4} \sin^2 \theta.$$

This is a symmetrical pressure distribution and leads to the theoretical results of zero lift and zero drag.

The theoretical result that drag is zero for all the above flows is called *d’Alembert’s paradox* and is a consequence of neglecting friction in the theory. In reality, skin friction and flow separation from the surface—both viscous effects—create a finite drag on any real aerodynamic body.

However, the theoretical result for lift obtained for the lifting cylinder is quite real, namely,

$$L = \rho_\infty V_\infty \Gamma.$$

This result is a general result for the inviscid incompressible flow over a cylindrical body of any arbitrary shape and is called the *Kutta-Joukowski theorem*. It states that the lift per unit span along the body is directly proportional to the circulation about the body. This result is the major focus of the circulation theory of lift, first developed in the early 1900s, and is still in use today for the prediction of the lifting characteristics of low-speed bodies.

For the source, vortex, and doublet flows described above, the center of the flow can be considered a *point* source, *point* vortex, or *point* doublet, where the point is

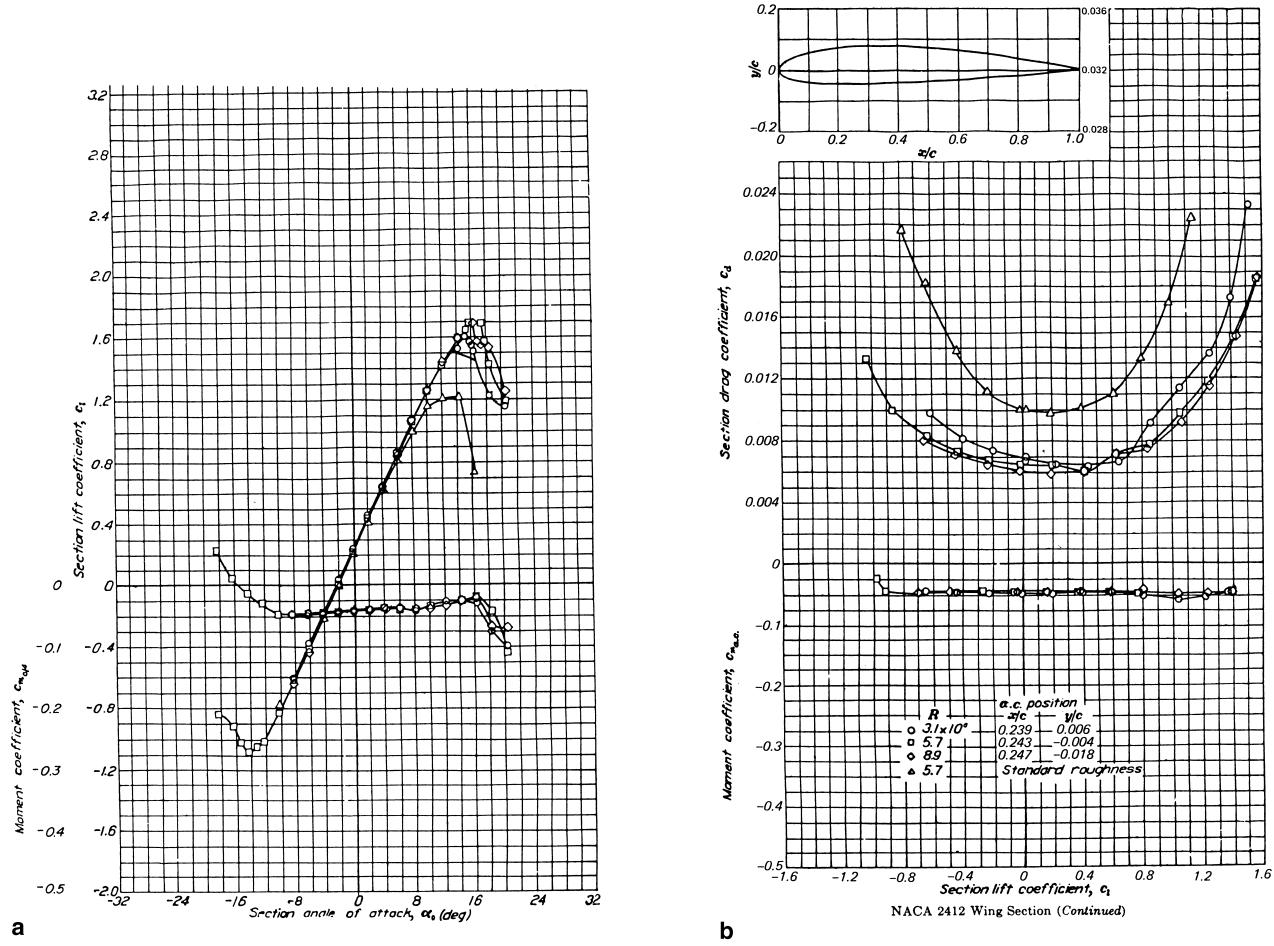


FIGURE 7 Lift and moment coefficients vs angle of attack for a conventional airfoil. (b) Drag and moment coefficients vs lift coefficient for a conventional airfoil.

a singularity since the velocity goes to infinity at the center. This leads to a general theoretical approach wherein the flow field over a body of general shape is calculated by distributing such singularities over the surface and calculating the strength of these singularities such that, in combination with the uniform free stream, the body shape becomes a streamline of the flow. Moreover, for an airfoil with a sharp trailing edge, the flow must leave smoothly at the trailing edge; this is in comparison to other possible solutions where it can leave the surface at some other point on the airfoil. Indeed, where the flow leaves the surface is a function of the amount of circulation Γ present around the airfoil. Nature chooses just the right value of Γ so that the flow always leaves smoothly at the sharp trailing edge. This is called the *Kutta condition*. If the airfoil is reasonably thin and the angle of attack is small, an approximate theory based on the idea of distributed singularities mentioned above (thin airfoil theory) leads to the following result for the rate of change of lift coefficient of c_1 vs angle of attack: $dc_1/d\alpha = 2\pi$ (per radian).

That is, for small angles of attack, the lift coefficient varies linearly with angle of attack—a result that is verified by experiment.

To further illustrate the aerodynamic properties of airfoils at low speed, Figs. 7a and 7b show experimental data for a National Advisory Committee for Aeronautics (NACA) 2412 airfoil; here, lift coefficient c_1 and moment coefficient about the quarter chord $c_{m_{1/4}}$ are shown as a function of angle of attack α (Fig. 7a), and the drag coefficient c_d and moment coefficient about the aerodynamic center $c_{m_{a.c.}}$ are shown as functions of c_1 (Fig. 7b). The aerodynamic center is defined as that point on the airfoil about which moments do not vary with angle of attack, as is clearly seen in Fig. 7b.

For three-dimensional bodies, the concept of distributed singularities over the body surface is also employed. Historically, for a finite wing, these singularities took the form of a number of vortex filaments (discrete lines of vorticity) in the shape of a horseshoe, with one segment running along the span of the wing (the so-called “lifting line”)

and the other segments trailing downstream into the wake of the wing. This lifting-line theory, due to Ludwig Prandtl (1875–1953), resulted in the first reasonable calculation of lift and drag on an airplane wing. It also identified a component of drag called *induced drag*, which is due to an alteration of the pressure distribution over the wing caused by strong vortices trailing downstream from the wing tips. These wing-tip vortices induce a general downward component of velocity over the wing (called down-wash), which in turn changes the pressure distribution in such a fashion as to increase the drag. This increase in drag is induced drag. Induced drag is directly proportional to the square of the lift coefficient; therefore, induced drag rapidly increases as the lift increases. For a complete airplane, the total drag coefficient C_D at low speeds can be expressed as

$$C_D = C_{D_0} + C_{D_i}, \quad (13)$$

where

$$C_{D_i} = C_L^2 / \pi AR.$$

In Eq. (13), which is called the *drag polar*, C_{D_0} is the parasite drag coefficient at zero lift; parasite drag is produced by the net effect of skin friction over the body surface plus the extra pressure drag produced by regions of flow separation over the surface (sometimes called form drag). Also in Eq. (13), C_{D_i} is the drag coefficient due to lift and is due to the increment of parasite drag (above the zero lift value) that occurs when the airplane is at the higher angles of attack associated with the production of lift, plus the induced drag due to the tip vortices, hence down-wash. In the above, AR is the aspect ratio, $AR = b^2/S$, where b is the wing span and S is the platform area of the wing. Also, e is an efficiency factor, where $e \leq 1$, and is, in part, associated with the mathematical shape of the spanwise distribution of lift over the wing.

In modern aerodynamics, the three-dimensional potential flow over low-speed bodies is calculated by means of singularities distributed over small, flat panels, and these panels in turn cover the surface of the body. The development of the “panel method,” using either, or a combination of, source, vortex, or doublet panels, has been well developed since 1965.

VI. INVISCID COMPRESSIBLE FLOW

Compressible flow differs from incompressible flow in at least the following respects:

1. The density of the flow becomes a variable.
2. The flow speeds are high enough that the flow kinetic energy becomes important and therefore energy

changes in the flow must be considered. This couples the science of thermodynamics into the aerodynamic considerations.

3. Shock waves can occur, which can completely dominate the flow.

The governing equations for inviscid, compressible flow are obtained from Eqs. (1)–(5) as

Continuity

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{V}) = 0 \quad (14)$$

Momentum (x component)

$$\rho \frac{Du}{Dt} = -\frac{\partial p}{\partial x} \quad (15)$$

Momentum (y component)

$$\rho \frac{Dv}{Dt} = -\frac{\partial p}{\partial y} \quad (16)$$

Momentum (z component)

$$\rho \frac{Dw}{Dt} = -\frac{\partial p}{\partial z}$$

Energy

$$\rho \frac{D(e + V^2/2)}{Dt} = p\dot{q} - \nabla \cdot (p\mathbf{V}) \quad (17)$$

Equations (14)–(17) are called the Euler equations. Note that for a compressible flow, Bernoulli’s equation [Eq. (7) or (8)] does *not* hold.

The speed of sound is an important parameter for compressible flow. The speed of sound a is given by $a^2 = (\partial p / \partial \rho)_s$, where the subscript indicates constant entropy. For a calorically perfect gas (a gas with constant specific heats and that obeys the gas law $p = \rho RT$),

$$a = \sqrt{\gamma RT},$$

where $\gamma = c_p/c_v$ and R is the specific gas constant. In turn, the speed of sound is used to define the *Mach number* as

$$M = V/a,$$

where V is the flow velocity.

Two defined properties particularly important to the analysis of compressible flow are (1) total temperature T_0 , defined as that temperature that would exist if the flow were brought to rest adiabatically, and (2) total pressure p_0 , defined as that pressure that would exist if the flow were brought to rest isentropically. For a calorically perfect gas, the relation between total and static properties is a function of γ and M only, as

$$\frac{T_0}{T} = 1 + \frac{\gamma - 1}{2} M^2$$

and

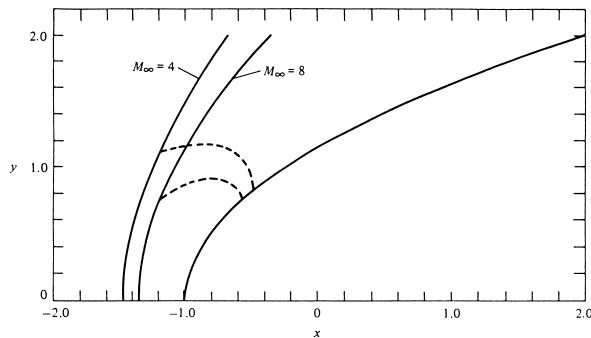


FIGURE 8 Supersonic flow over a blunt body; shock waves and sonic lines for two different freestream Mach numbers. [From Anderson, J. D., Jr. (2001). “Fundamentals of Aerodynamics,” 3rd ed., McGraw-Hill, New York.]

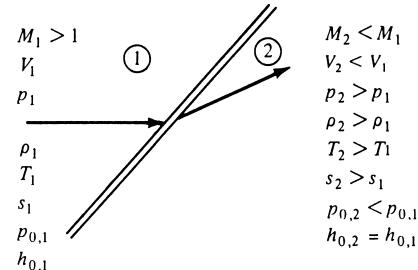
$$\frac{p_0}{p} = \left(1 + \frac{\gamma - 1}{2} M^2\right)^{\gamma/(\gamma-1)}. \quad (18)$$

Eq. (18) is particularly useful for the measurement of Mach number in subsonic compressible flow. A Pitot tube in the flow will sense p_0 ; if the static pressure p is known at the same point, then the Pitot measurement of P_0 will directly yield the Mach number via Eq. (18).

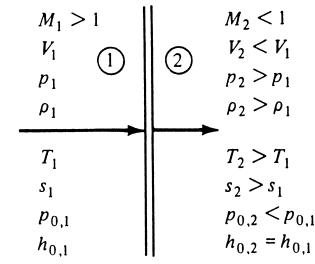
In any steady flow where $M > 1$, shock waves can occur. For example, consider the blunt-nosed, parabolically shaped body shown at the right in Fig. 8. Flow is moving from left to right. If the upstream flow is supersonic ($M_\infty > 1$), then a curved shock wave will exist slightly upstream of the blunt nose. Two such shocks are shown to the left of the body; the leftmost shock illustrates the case for $M_\infty = 4$, and the rightmost shock pertains to the case for $M_\infty = 8$. The shock waves slow down the flow. The lower part of the flow field behind the shock wave is locally subsonic, and the upper part, after sufficient expansion around the body, is locally supersonic, albeit at a Mach number lower than M_∞ . The dashed lines shown in Fig. 8 divide the subsonic and supersonic regions behind the shock and are called the *sonic lines*, since the local Mach number $M = 1$ along these lines. Note from Fig. 8 that as the freestream Mach number increases, the shock wave moves closer to the body and the sonic line moves down closer to the centerline of the flow. An essential ingredient of the understanding of supersonic flow is the calculation of the shape and strength of shock waves, such as those illustrated in Fig. 8. Therefore, let us examine shock waves in more detail.

A shock wave is an extremely thin region, typically on the order of 10^{-5} cm, across which the flow properties can change drastically. The shock wave is usually at an oblique angle to the flow, such as sketched in Fig. 9a; however, there are many cases where we are interested in a shock

wave normal to the flow, as sketched in Fig. 9b. For example, referring again to Fig. 8, the portion of the bow shock wave directly in front of the nose is normal, whereas the shock wave is oblique away from the nose. In both cases of a normal or an oblique shock, the shock wave is an almost explosive compression process, where the pressure increases almost discontinuously across the wave. As shown in Fig. 9, in region 1 ahead of the shock, the Mach number, flow velocity, pressure density, temperature, entropy, total pressure, and total enthalpy (defined as $e + p/\rho + V^2/2$) are denoted by M_1 , V_1 , ρ_1 , T_1 , s_1 , $p_{0,1}$, and $h_{0,1}$, respectively. The analogous quantities in region 2 behind the shock are denoted by a subscript 2. The qualitative changes across the wave are noted in Fig. 9. The pressure, density, temperature, and entropy increase across the shock, whereas the total pressure, Mach number, and velocity decrease. Physically, the flow across a shock wave is adiabatic, which leads to a constant total enthalpy across the wave. For a perfect gas, this also means that the total temperature is constant across the shock, i.e., $T_{0,1} = T_{0,2}$. In both oblique shock and normal shock cases, the flow ahead of the shock must be supersonic, that is, $M_1 > 1$. Behind the oblique shock, the flow usually remains supersonic, that is, $M_2 > 1$, but at a reduced Mach number,



(a) Oblique shock wave



(b) Normal shock wave

FIGURE 9 Oblique and normal shock waves. [From Anderson, J. D., Jr. (2001). “Fundamentals of Aerodynamics,” 3rd ed., McGraw-Hill, New York.]

$M_2 < M_1$. However, there are a few special cases where the oblique shock is strong enough to decelerate the downstream flow to a subsonic Mach number. For the normal shock, as sketched in Fig. 9b, the downstream flow is always subsonic, that is, $M_2 < 1$.

The quantitative changes across a normal shock wave can be obtained from Eqs. (14)–(17) specialized for steady one-dimensional flow and integrated to obtain the following basic normal shock equations:

Continuity

$$\rho_1 V_1 = \rho_2 V_2$$

Momentum

$$p_1 + \rho_1 V_1^2 = p_2 + \rho_2 V_2^2$$

Energy

$$e_1 + p_1/\rho_1 + V_1^2/2 = e_2 + p_2/\rho_2 + V_2^2/2$$

Along with the equations of state $p = \rho RT$ and $e = c_v T = RT/(\gamma - 1)$ for a calorically perfect gas, the basic normal shock equations can be algebraically manipulated to obtain the following changes of flow properties across the normal shock:

$$\frac{p_2}{p_1} = 1 + \frac{2\gamma}{\gamma + 1}(M_1^2 - 1) \quad (19)$$

$$\frac{\rho_2}{\rho_1} = \frac{V_1}{V_2} = \frac{(\gamma + 1)M_1^2}{2 + (\gamma - 1)M_1^2} \quad (20)$$

$$M_2^2 = \frac{1 + [(\gamma - 1)/2]M_1^2}{\gamma M_1^2 - (\gamma - 1)/2} \quad (21)$$

Note that the changes across a normal shock wave depend only on the value of γ and the upstream Mach number M_1 .

For an oblique shock wave, let β be the angle between the shock wave and the upstream flow direction: β is called the wave angle. Also, note in Fig. 9a that the flow, in crossing the oblique shock, is bent in the upward direction behind the shock. Let θ be the angle between the downstream and upstream flow directions: θ is called the deflection angle. Then, Eqs. (19)–(21) hold for an oblique shock if M_1 , V_1 , M_2 , and V_2 are replaced by $M_1 \sin \beta$, $V_1 \sin \beta$, $M_2 \sin(\beta - \theta)$, and $V_2 \sin(\beta - \theta)$, respectively. Moreover, for an oblique shock, the wave angle, deflection angle, and upstream Mach number are related through the equation

$$\tan \theta = 2 \cot \beta \frac{M_1^2 \sin^2 \beta - 1}{M_2^2(\gamma + \cos 2\beta) + 2}. \quad (22)$$

From Eqs. (19)–(21) written appropriately for an oblique shock, combined with Eq. (22), we see that if we know any two quantities about the shock, say β and M_1 , all

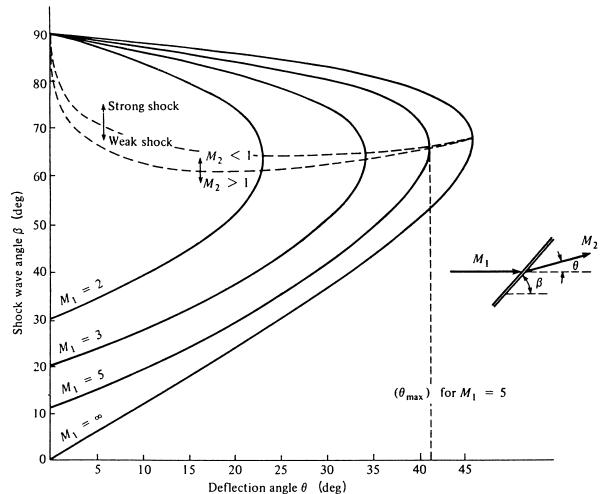


FIGURE 10 Oblique shock wave relations. [From Anderson, J. D., Jr. (1990). "Modern Compressible Flow," 2nd ed., McGraw-Hill, New York.]

other quantities such as θ , M_2 , p_2/p_1 , etc. are determined. A plot of Eq. (22) for $\gamma = 1.4$ (air at standard conditions) is given in Fig. 10. From this figure, we can deduce the following physical characteristics about oblique shock waves:

1. For any given M_1 , there is a maximum deflection angle θ_{\max} . If the physical geometry is such that $\theta > \theta_{\max}$, then no solution exists for a straight oblique shock wave; in such a case, the flow field will adjust in such a fashion to curve and detach the shock wave.

2. For any given $\theta < \theta_{\max}$, there are two values of β predicted by Fig. 10 for a given M_1 . The larger value of β is called the strong shock solution, and the smaller β is the weak shock solution. In nature, the weak shock solution is favored and is the one that usually occurs.

3. The downstream Mach number M_2 is subsonic for the strong shock solutions, whereas M_2 is supersonic for the weak shock solutions, except for the narrow band between the two dashed lines in Fig. 10, where $M_2 < 1$ even for the weak oblique shock case.

A Pitot tube can be used to measure the Mach number in a supersonic flow. However, in contrast to our previous discussions of the use of a Pitot tube, for the supersonic flow case a normal shock wave will exist in front of the mouth of the tube. Hence, the Pitot tube will measure the total pressure behind the normal shock, *not* the total pressure of the flow itself as in the subsonic case. By appropriate manipulation of the basic normal shock equations, the following relation can be found for a Pitot tube in supersonic flow:

$$\frac{p_{0.2}}{p_1} = \left[\frac{(\gamma + 1)^2 M_1^2}{4\gamma M_1^2 - 2(\gamma - 1)} \right]^{\gamma/(\gamma-1)} \left(\frac{1 - \gamma + 2\gamma M_1^2}{\gamma + 1} \right), \quad (23)$$

where $p_{0.2}$ is the total pressure behind the normal shock (the pressure measured by the tube), p_1 is the static pressure upstream of the normal shock, and M_1 is the Mach number upstream of the shock. If a Pitot tube measurement is taken at a given point in a supersonic flow, and if the static pressure is known at that point, then the Pitot measurement will directly yield the local Mach number at that point via Eq. (23).

An oblique shock, in the limit of infinitely weak strength, becomes a *Mach wave*, where $\theta = 0$, and $\beta = \mu$, where μ is called the *Mach angle*:

$$\mu = \sin^{-1}(1/M).$$

For all oblique shocks of a finite strength, $\beta > \mu$.

Oblique shock waves are created when a supersonic flow is bent into itself, such as the flow over a concave corner as shown in Fig. 11a. In contrast, when a supersonic flow is bent away from itself, such as the flow over a convex corner as sketched in Fig. 11b, an *expansion wave* is created. Expansion waves are the opposite of shock waves. Expansions are composed of an infinite number of Mach waves and are a region of smooth continuous change through the expansion fan. Across the expansion wave, the Mach number increases, and the pressure, temperature, and density decrease. The flow through an expansion wave is isentropic (constant entropy), and hence, both the total pressure and the total temperature are constant through the wave. For a given Mach number M_1 ahead of the expansion wave and a given deflection angle θ , the Mach number M_2 behind the expansion can be found from

$$\theta = v(M_2) - v(M_1),$$

where v is the *Prandtl–Meyer function*, given by

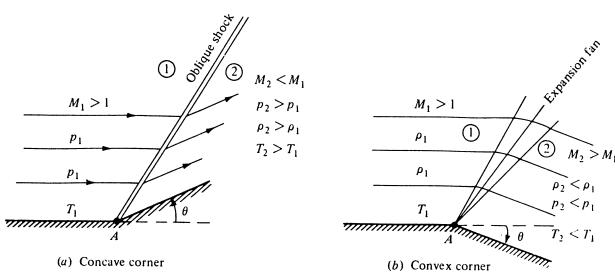


FIGURE 11 Oblique shock and expansion waves. [From Anderson, J. D., Jr. (2001). “Fundamentals of Aerodynamics,” 3rd ed., McGraw-Hill, New York.]

$$v(M) = \sqrt{\frac{\gamma+1}{\gamma-1}} \tan^{-1} \sqrt{\frac{\gamma-1}{\gamma+1} (M^2 - 1)} - \tan^{-1} \sqrt{M^2 - 1}$$

For a calorically perfect gas. The corresponding temperature and pressure changes can be obtained from

$$\frac{T_2}{T_1} = \frac{1 + [(\gamma - 1)/2]M_1^2}{1 + [(\gamma - 1)/2]M_2^2}$$

and

$$\frac{p_2}{p_1} = \left(\frac{T_2}{T_1} \right)^{\gamma/(\gamma-1)}.$$

The external supersonic flow fields over some aerodynamic bodies can sometimes be synthesized in an approximate fashion by a proper combination of local oblique shock and expansion wave solutions. Such approaches are called *shock-expansion solutions*.

Compressible flows through ducts (i.e., internal compressible flows) are of great importance in the design of high-speed wind tunnels, jet engines, and rocket engines, to name just a few applications. Consider a duct with a local cross-sectional area A . The area may change with length along the duct. By a proper combination of the continuity, momentum, and energy equations for a compressible flow, the following relation can be extracted, which governs the compressible flow in a variable area duct, where the flow properties are assumed to be uniform across any cross section but can change from one cross section to another (so-called quasi-one-dimensional flow):

$$\frac{dA}{A} = (M^2 - 1) \frac{dV}{V}. \quad (24)$$

In Eq. (24), dA is the local change in area, dV is the corresponding change in velocity, and M is the local Mach number. From Eq. (24), we see that

1. For *subsonic flow* ($0 \leq M < 1$), V increases as A decreases, and V decreases as A increases. Therefore, to increase the velocity, a *convergent* duct must be used, whereas to decrease the velocity, a *divergent* duct must be employed.

2. For *supersonic flow* ($M > 1$), V increases as A increases, and V decreases as A decreases. Hence, to increase the velocity, a *divergent* duct must be used, whereas to decrease the velocity, a *convergent* duct must be employed.

3. When $M = 1$, $dA = 0$. Hence, sonic flow will occur in that location inside a variable-area duct where the area variation is a local minimum. Such a location is called a *sonic throat*.

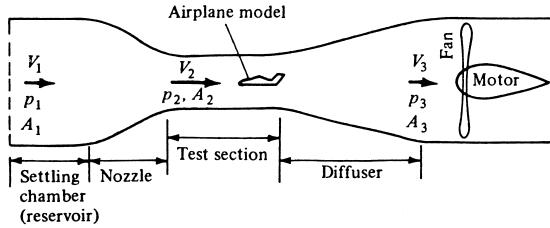


FIGURE 12 Sketch of a subsonic wind tunnel. [From Anderson, J. D., Jr. (2001). “Fundamentals of Aerodynamics,” 3rd ed., McGraw-Hill, New York.]

Clearly, subsonic and supersonic flows through a changing-area duct behave differently. This is why subsonic wind tunnels are configured differently than supersonic wind tunnels. A sketch of a basic subsonic tunnel is shown in Fig. 12. Here, we see a convergent duct for speeding up the subsonic flow (the nozzle), a constant-area section where a test model is placed in the high-speed (but subsonic) flow, and then a divergent duct (the diffuser) for slowing the flow down before exhausting it to the surroundings. In contrast, a sketch of a basic supersonic tunnel is shown in Fig. 13. Here, the initially low-speed subsonic flow from a reservoir is speeded up in a convergent section, reaching Mach 1 at the throat, and then the now-supersonic flow is further speeded up in a divergent section. Hence, a nozzle designed to produce a supersonic flow is a *convergent-divergent* nozzle, sometimes called a *de Laval nozzle*, after Carl G. P. de Laval (1845–1912), who first employed such nozzles in a steam turbine. The high-speed supersonic flow then enters a constant-area test section where a test model is placed. Downstream of the model, the flow, along with the shock wave system produced by the model, enters a diffuser designed to slow the flow to a low subsonic speed before exhausting to the surroundings. To accomplish this, the initially supersonic flow must first be slowed in a convergent section to Mach 1 at a minimum area (the “second” throat) and then further slowed subsonically in a divergent section. Hence, the supersonic diffuser is also a convergent–divergent duct, just as in the case of the supersonic nozzle; however, their functions are completely opposite.

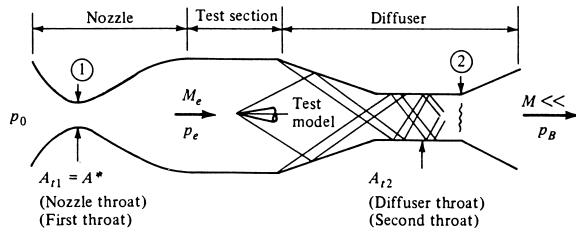


FIGURE 13 Sketch of a supersonic wind tunnel. [From Anderson, J. D., Jr. (2001). “Fundamentals of Aerodynamics,” 3rd ed., McGraw-Hill, New York.]

The flow in a properly designed supersonic nozzle is isentropic. For this case, and for a calorically perfect gas, the Mach number that exists at a given cross section with area A is simply a function of the ratio of this area to the throat area, denoted by A^* . That is,

$$\left(\frac{A}{A^*}\right)^2 = \frac{1}{M^2} \left[\frac{2}{\gamma + 1} \left(1 + \frac{\gamma - 1}{2} M^2 \right) \right]^{(\gamma+1)/(\gamma-1)}$$

This is a double-valued function; for a given A/A^* , two solutions exist for M —a subsonic value in the convergent portion of the nozzle and a supersonic value in the divergent portion.

Let us return to the consideration of external compressible flows. If the flow is subsonic or supersonic, irrotational, and involves the flow over slender bodies at small angle of attack, Eqs. (14)–(17) reduce to a single linearized equation:

$$(1 - M_\infty^2) \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} = 0, \quad (25)$$

where ϕ is a *perturbation* velocity potential defined such that $\partial \phi / \partial x = u$, $\partial \phi / \partial y = v$, and $\partial \phi / \partial z = w$, where u , v , and w are small perturbation velocities superimposed on the uniform free stream in such a manner that $u = V_\infty + u$, $v = v$, and $w = w$. Equation (25) does *not* hold for transonic or hypersonic Mach numbers.

A solution of Eq. (25) for a subsonic flow leads to a method for correcting low-speed, incompressible flow data to take into account compressibility effects—so-called *compressibility corrections*. For example, if C_{p0} denotes the low-speed incompressible flow value of the pressure coefficient at a certain point on a two-dimensional airfoil, then the value of C_p at the same point for a high-speed subsonic freestream Mach number M_∞ is given approximately by $C_p = C_{p0} / \sqrt{1 - M_\infty^2}$. This compressibility correction is called the *Prandtl–Glauert rule* and is generally valid for $M_\infty < 0.7$. In turn, the lift and moment coefficients for the airfoil are similarly related to their corresponding low-speed incompressible values c_{l0} and c_{m0} as

$$c_l = \frac{c_{l0}}{\sqrt{1 - M_\infty^2}} \quad c_m = \frac{c_{m0}}{\sqrt{1 - M_\infty^2}}.$$

Equation (25) can also be solved for supersonic flows, leading to a simple expression for the pressure coefficient for two-dimensional flow:

$$C_p = \frac{2\theta}{\sqrt{M_\infty^2 - 1}}, \quad (26)$$

where θ is the local angle between the tangent to the body and the freestream direction and M_∞ is the freestream Mach number. In turn, this result from linearized theory can be used to calculate the lift and wave drag coefficients

for thin airfoils at a small angle of attack. For example, for a flat plate at angle of attack α , Eq. (26) leads to

$$c_l = \frac{4\alpha}{\sqrt{M_\infty^2 - 1}}, \quad (27)$$

$$c_d = \frac{4\alpha^2}{\sqrt{M_\infty^2 - 1}}. \quad (28)$$

The wave drag coefficient c_d is a ramification of the high pressure behind a shock wave and can be associated with the entropy increase across a shock wave. We can view wave drag as that component of drag that is associated with shock waves on the body; obviously, wave drag exists only at supersonic and hypersonic speeds. For thin airfoils at small α , Eq. (27) still holds; however, Eq. (28) is modified to take into account the finite thickness and curvature (camber) of the airfoil.

Modern exact calculations of inviscid compressible flows—subsonic, transonic, supersonic, and hypersonic—are made by means of sophisticated numerical solutions of Eqs. (14)–(17) on a high-speed digital computer.

As a final note in this section, a comment is made about hypersonic flow. Such flow, which by rule of thumb is denoted as the regime where $M_\infty > 5$, is characterized by several important physical phenomena already described in Section IV. Among these is the existence of thin shock layers around the body (the shock waves lie close to the body). For such a flow, the physical flow picture looks very much like the original model proposed by Newton in 1687 (as described in Section I). Hence, Newton's sine-squared law is a reasonable prediction for the pressure coefficient at hypersonic speeds, that is,

$$C_p = 2 \sin^2 \theta,$$

where θ is the angle between a tangent to the surface and the freestream direction. A *modified* Newtonian formula actually gives improved results; that is,

$$C_p = C_{p_{\max}} \sin^2 \theta,$$

where $C_{p_{\max}}$ is the pressure coefficient at the stagnation point after the flow has passed through a normal shock wave at the freestream Mach number.

VII. VISCOUS FLOW

The aerodynamic results discussed in the previous sections dealt with inviscid flows. In contrast, in this section we examine the impact of transport phenomena that exert dissipative effects in a flow. In particular, we examine the influence of viscosity and thermal conduction; mass diffusion will not be considered here.

In the governing equations for fluid dynamics, Eqs. (1)–(5), the viscous effects are present through the shear stress, normal stress, and heat conduction terms. Shear stress and normal stress are directly proportional to the velocity gradients in the flow, for example, $\tau_{yx} = \mu(\partial u / \partial y)$; hence, viscosity effects are large in regions of large velocity gradients. Similarly, thermal conduction involves the transport of heat in the opposite direction of a temperature gradient, for example, $\dot{q}_y = -k(\partial T / \partial y)$, where \dot{q}_y is the heat flux (energy per unit area per unit time) in the y direction, k is the thermal conductivity, and $\partial T / \partial y$ is the temperature gradient in the y direction. Therefore, thermal conduction effects are large in regions of large temperature gradients.

The major practical effects of viscous flow on aerodynamic problems are the following:

1. The action of viscosity generates a shear stress at a solid surface, which in turn creates a drag called skin friction drag.

2. Shear stress acting on a surface tends to slow the flow velocity near the surface. If the flow is experiencing an adverse pressure gradient (a region where the pressure increases in the flow direction), then the low-energy fluid elements near the surface cannot negotiate the adverse pressure gradient; as a result, the flow separates from the surface. Flow separation alters the pressure distribution over the surface in such a fashion to increase the drag; this is called “pressure drag due to flow separation,” or “form drag.” In addition, if the body is producing lift, then flow separation can greatly reduce the lift. This is the mechanism that limits the lift coefficient on an airfoil, wing, or lifting body to some maximum value. For example, returning to Fig. 7a, note that c_l increases with α until a maximum value is achieved. As α is further increased, massive flow separation occurs, which causes the lift to rapidly decrease. Under this condition, the airfoil is said to be “stalled.”

3. The kinetic energy of the fluid elements near the surface is reduced due to the retarding effects of friction; in turn, this energy is converted to thermal energy in the flow near the surface, thus increasing the temperature of the flow. For high-speed flows, particularly at high supersonic and hypersonic speeds, this dissipative phenomenon can create very high temperatures near the surface. Through the mechanism of thermal conduction at the surface, large aerodynamic heating rates can result.

The magnitude of the skin friction and aerodynamic heating and the extent of flow separation are greatly influenced by the nature of the viscous flow; that is, whether the flow is laminar or turbulent. In a laminar flow, the path lines of various fluid elements are smooth and regular. In

contrast, if the motion of a fluid element is very irregular and tortuous, the flow is called turbulent flow. The net effect of turbulence is to pump some of the higher energy flow that exists far away from the surface to a region closer to the surface. This increases the velocity and temperature gradients at the surface, hence increasing skin friction and heat transfer. From this point of view, turbulent flow is a detriment. On the other hand, flow separation is delayed by turbulent flow, hence reducing the pressure drag due to flow separation. In this fashion, turbulent flow is advantageous. In most aerodynamic problems, the viscous flow first begins as laminar; then at some downstream location the flow experiences a transition to turbulent. There is a basic principle that nature always moves toward the state of maximum disorder; in terms of viscous flow, this means that nature is always driving toward turbulent flow. The location of the transition region is influenced by various phenomena. For example, if the Reynolds number Re_∞ is increased, transition moves forward. If the Mach number M_∞ is increased, transition moves rearward. If the surface is roughened, transition moves forward. If the surface is highly cooled, transition moves rearward. If the amount of natural turbulence inherently present in the free stream is increased, transition moves forward. These are just a few of the many phenomena that influence transition from laminar to turbulent flow.

Viscous flow problems are most exactly analyzed by solving the complete Navier–Stokes equations. Eqs. (1)–(5). However, these nonlinear equations are so complex that no general analytical solution exists, and exact solutions for practical aerodynamic applications are only presently being obtained through detailed finite-difference numerical solutions on a high-speed computer. Even in these cases, some Navier–Stokes solutions require many hours of computer time to go to completion. This is a current state-of-the-art problem in the analysis of viscous flow.

Faced with these difficulties in solving Eqs. (1)–(5), Ludwig Prandtl in 1904 introduced the *boundary layer concept*. Referring again to Fig. 1, in many problems most of the viscous effects are limited to a thin region adjacent to the solid surface. This thin viscous region is called the *boundary layer*. For such a region, the governing equations, Eqs. (1)–(5), can be reduced to the *boundary layer equations*, which in two-dimensional flow are

Continuity

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

x Momentum

$$\rho u \frac{\partial u}{\partial x} + \rho v \frac{\partial u}{\partial y} = -\frac{dp_e}{dx} + \frac{\partial}{\partial y} \left(\mu \frac{\partial u}{\partial y} \right)$$

y Momentum

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

Energy

$$\rho u \frac{\partial h}{\partial x} + \rho v \frac{\partial h}{\partial y} = \frac{\partial}{\partial y} \left(k \frac{\partial T}{\partial y} \right) + u \frac{dp_e}{dx} + \mu \left(\frac{\partial u}{\partial y} \right)^2$$

where x and y are coordinates tangential and normal to the surface respectively. Since $\partial p/\partial y = 0$ through the boundary layer, then the pressure distribution at the outer edge of the boundary layer, $p_e = f(x)$, is impressed directly through the boundary layer to the surface.

The boundary layer equations can be readily solved for flow over a flat plate. Classical solutions for incompressible laminar flow yield

$$c_f = \frac{0.664}{\sqrt{Re_x}}$$

and

$$\delta = \frac{5.0x}{\sqrt{Re_x}},$$

where c_f is the skin friction coefficient defined as $c_f = \tau_w / (\frac{1}{2} \rho_\infty V_\infty^2)$, Re_x is the Reynolds number based on running length from the leading edge, $Re_x \equiv \rho_\infty V_\infty x / \mu_\infty$, and δ is the thickness of the boundary layer. Note from these results that, for a laminar boundary layer, c_f and hence τ_w varies as $x^{-1/2}$, and δ grows as \sqrt{x} .

Results for a laminar compressible flow over a flat plate indicate that c_f and δ are functions not only of Re_x but of M_∞ , Pr_∞ , and T_w/T_∞ as well, where Pr_∞ is the freestream Prandtl number, T_w is the surface temperature, and T_∞ is the freestream temperature. We can write, for compressible laminar flow,

$$c_f = \frac{0.664}{\sqrt{Re_x}} f(M_\infty, Pr_\infty, T_w/T_\infty)$$

and

$$\delta = \frac{5.0}{\sqrt{Re_x}} g(M_\infty, Pr_\infty, T_w/T_\infty).$$

Results for incompressible turbulent flow cannot be obtained exactly from the boundary layer equations. Assumptions must be made about a model for the turbulence, and empirical data is always needed to some stage in a turbulent-flow analysis. However, approximate results for a flat plate are

$$c_f = \frac{0.058}{Re_x^{0.2}}$$

and

$$\delta = \frac{0.37x}{Re_x^{0.2}}.$$

Here, c_f is seen to vary as $x^{-0.2}$, and the turbulent boundary layer thickness grows as $x^{0.8}$. As in the case of laminar flow, these incompressible flow results become modified for compressible flow, where c_f and δ are both functions of M_∞ , Pr_∞ , and T_w/T_∞ , in addition to Re_x .

VIII. TWO MODERN AERODYNAMIC INVENTIONS

The drag coefficient for an airfoil as a function of Mach number at subsonic speeds is sketched qualitatively in Fig. 14. Over a large part of the Mach number range, c_d is relatively constant. This pertains to the completely subsonic flow over the airfoil, as sketched in Fig. 4a. As the freestream Mach number is increased, the flow Mach number on the top surface of the airfoil also increases. That freestream Mach number at which sonic flow is first achieved somewhere on the airfoil surface is called the *critical Mach number*, denoted by M_{cr} in Fig. 14. If M_∞ is further increased, an entire pocket of supersonic flow develops on the top (and even possibly the bottom) surface of the airfoil, as sketched in Fig. 4b. This corresponds to point b in Fig. 14. With the appearance of the shock waves at the end of these supersonic pockets, the drag coefficient begins to rapidly increase. The freestream Mach number at which this major drag increase begins is called the *drag-divergence Mach number*, also shown in Fig. 14. The major increase in drag near a freestream Mach number of 1 was once thought to be so severe that airplanes would never fly faster than sound. This myth of the “sound barrier” was supported by theoretical results such as Eq. (28), which shows c_d going to infinity as M_∞ goes to 1. How-

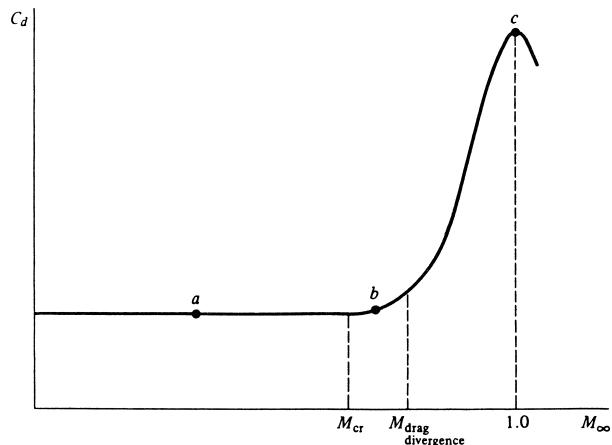


FIGURE 14 Schematic of the variation of drag coefficient for an airfoil as a function of freestream Mach number at subsonic and transonic speeds. [From Anderson, J. D., Jr. (2000). “Introduction to Flight,” 4th ed., McGraw-Hill, New York.]

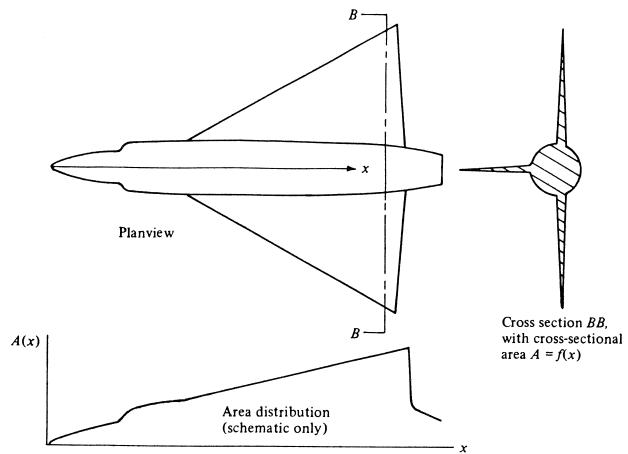


FIGURE 15 Example of a non-area-ruled airplane. [From Anderson, J. D., Jr. (2001). “Fundamentals of Aerodynamics,” 3rd ed., McGraw-Hill, New York.]

ever, the linear theory that produces this result is not valid near Mach 1. In reality, the drag coefficient will peak at some finite value around Mach 1, as shown in Fig. 14; as long as the airplane has excess thrust from the engines to overcome this peak drag, the aircraft can easily fly into the supersonic regime.

Two major practical aerodynamic inventions since 1950 have been (1) the area rule and (2) the supercritical airfoil. The former is an effort to reduce the peak value of c_d shown in Fig. 14, and the latter attempts to push the drag-divergence Mach number to a higher value.

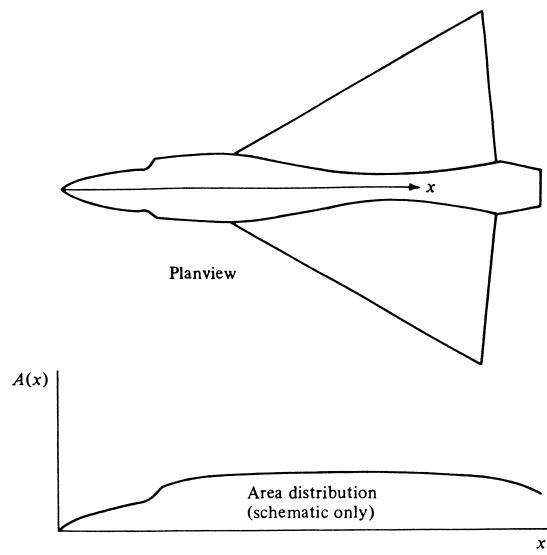


FIGURE 16 Example of an area-ruled airplane. [From Anderson, J. D., Jr. (2001). “Fundamentals of Aerodynamics,” 3rd ed., McGraw-Hill, New York.]

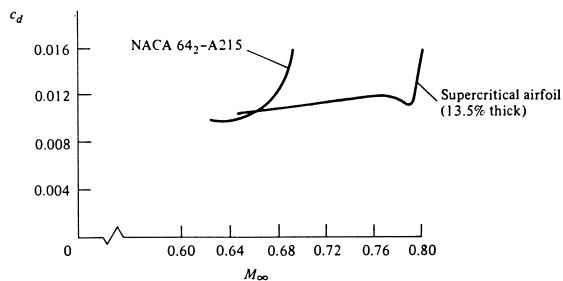


FIGURE 17 Effect of a supercritical airfoil on the drag-divergence Mach number. [From Andeson, J. D., Jr. (2001). “Fundamentals of Aerodynamics,” 3rd ed., McGraw-Hill, New York.]

The area rule states that the distribution of cross-sectional area of the body should be smooth and gradual. This implies that the fuselage area must be decreased in the vicinity of the wing. Qualitative sketches of non-area-ruled and area-ruled aircraft are shown in Figs. 15 and 16, respectively. The peak transonic drag coefficient is less for the area-ruled airplane in Fig. 16 than for the non-area-ruled aircraft in Fig. 15. Thus, the area rule allows high-performance airplanes to routinely break the speed of sound.

The supercritical airfoil is designed to avoid the large pocket of supersonic flow over the airfoil until larger values of M_∞ are achieved. In other words, the supercritical airfoil increases the drag-divergence Mach number, as shown in Fig. 17.

The development of both the area rule in the 1950s and the supercritical airfoil in the 1960s has greatly increased the speed and efficiency of transonic and supersonic airplanes. It is noteworthy that both of these major developments stemmed from the same person. Richard T. Whitcomb (1921–), who was an aerodynamist for the National Advisory Committee for Aeronautics and later for the National Aeronautics and Space Administration (NASA).

IX. EPILOGUE

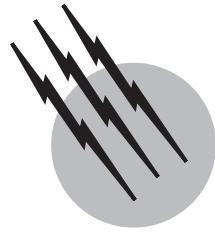
This short article on aerodynamics has attempted to present a bird’s-eye view of the subject. Many details and other important subjects are not discussed. Therefore, the reader is encouraged to examine the complete books on aerodynamics listed in the bibliography for a considerable amplification of material on the subject.

SEE ALSO THE FOLLOWING ARTICLES

AIRCRAFT AERODYNAMIC BOUNDARY LAYERS • AIRCRAFT PERFORMANCE AND DESIGN • AIRCRAFT SPEED AND ALTITUDE • COMPUTATIONAL AERODYNAMICS • FLOW VISUALIZATION • FLUID DYNAMICS • HEAT TRANSFER • JET AND GAS TURBINE ENGINES

BIBLIOGRAPHY

- Anderson, J. D., Jr. (1990). “Modern Compressible Flow; with Historical Perspective,” 2nd ed., McGraw-Hill, New York.
- Anderson, J. D., Jr. (2000). “Introduction to Flight,” 4th ed., McGraw-Hill, New York.
- Anderson, J. D., Jr. (2001). “Fundamentals of Aerodynamics,” 3rd ed., McGraw-Hill, New York.
- Bertin, J. J., and Smith, M. L. (1979). “Aerodynamics for Engineers,” Prentice Hall, Englewood Cliffs, NJ.
- Katz, J., and Plotkin, A. (1991). “Low-Speed Aerodynamics: From Wing Theory to Panel Methods,” McGraw-Hill, New York.
- Kuethe, A. M., and Chow, C. Y. (1986). “Foundations of Aerodynamics,” 4th ed., Wiley, New York.
- Marthy, T., and Brebbia, C. (1990). “Computational Methods in Viscous Aerodynamics,” Computational Mechanics, Southampton.
- McCormick, B. W. (1979). “Aerodynamics, Aeronautics, and Flight Mechanics,” Wiley, New York.
- Shivell, R. S. (1983). “Fundamentals of Flight,” Prentice-Hall, Englewood Cliffs, NJ.



Helicopters

Barnes W. McCormick

Pennsylvania State University

I. History

II. Technical Considerations

GLOSSARY

Articulated rotor One that is hinged to allow flapping of the rotor (also, lead-lag motion).

Collective pitch Average pitch angle for all of the rotor pitch blades from which the cyclic pitch angle is measured.

Cyclic pitch Periodic variation in the pitch angle as the rotor blade rotates.

Feathering Rotation of the blade about a radialwise axis along the blade that changes its pitch angle.

Flapping Unsteady motion of the blade moving in and out of the disk or shaft plane (out-of-plane motion).

Hingeless rotor Blade without hinges that flaps due to its flexibility; sometimes called by the misnomer *rigid rotor*.

In-ground effect (IGE) Usually taken to mean that a helicopter is hovering within one rotor diameter above the ground, so that the rotor aerodynamics are affected by the presence of the ground.

Lead lag Unsteady motion fore and aft relative to the blade and superimposed on its steady rotation (in-plane motion).

Out-of-ground effect (OGE) Hovering far enough above the ground, usually more than one rotor diameter, so that the rotor aerodynamics are not affected by the ground.

Pitch angle Angle between the rotor blade and the shaft plane.

Shaft plane Plane perpendicular to rotor shaft (also called the disk plane).

Teetering rotor Two-bladed rotor consisting of a continuous structure that teeters on the hub in a manner similar to that of a seesaw.

A **HELICOPTER** is an aircraft that can take off vertically and hover without motion in the air. It is capable of limited rearward and sideward flight and can achieve forward flight speeds of up to ~ 200 mph depending on the particular design. To achieve its very low speed and hovering performance, the helicopter derives its lift from one or more relatively large rotors that rotate on a powered shaft. Thus, an airflow is created over the rotor blades independent of forward speed, which results in lift being produced on the blades.

I. HISTORY

It is difficult to identify any one person as the inventor of the helicopter or to say that on a specific date the first successful flight of a helicopter was accomplished. Rather, the helicopter evolved and was refined over a period of

approximately 35 years, culminating in the development of the main rotor-tail rotor configuration conceived by Igor Sikorsky. His VS-300 was first flown in its final configuration on December 8, 1941. Before that date, it had been flown with three tail rotors to provide control and balance the torque of the main rotor. Sikorsky's development of the single rotor-tail rotor configuration, undoubtedly a major accomplishment in the development of the helicopter as we know it today, was nevertheless only one of the major milestones in the evolution of the helicopter. It was the accomplishments of others that led Sikorsky to his own significant achievement. The serious engineering efforts to develop a human-carrying machine followed models and toys, such as the Chinese top, dating back to the fourteenth century. One of Leonardo da Vinci's notebooks, dated 1480, shows a model helicopter driven by a clockwork motor. The rotor is a continuous helical surface making approximately one turn. His notes imply that the model flew, but from his sketch an antitorque device is not apparent. In 1853, Sir George Cayley, of fixed-wing fame, designed a Chinese top that had three pitched blades resembling the rotor system of a modern helicopter. It was not until a half-century later that a person was finally lifted vertically from the ground by means of power. Chronologically, significant milestones in the development of the helicopter are listed below.

September 29, 1907. The Brequet-Richet Gyroplane No. 1, designed by Charles Brequet, first flew in Douai, France. Lifted by four 26-ft-diameter rotors, this fragile craft hovered ~2 ft off the ground for a minute while being restrained by four men holding onto the frame. Although without a control system it was far from being a practical helicopter, the craft nevertheless demonstrated that it was possible to lift a person vertically by means of powered rotors.

November 13, 1907. Frenchman Paul Carnu became the first person in history to rise vertically, completely unrestrained by any support. [Figure 1](#) is a photograph of Carnu's helicopter. Two rotors were attached to each end of a skeletal frame. A 24-hp engine turned the rotors at

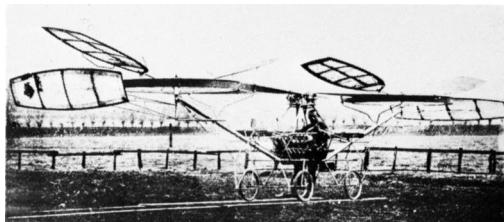


FIGURE 1 Carnu helicopter. Frenchman Paul Carnu was the first person to rise vertically unrestrained on November 13, 1907. [Courtesy of The Smithsonian Institution.]

~90 rpm. Carnu was possibly the first experimenter to be concerned with stability and control. He mounted, just below and outside of each rotor, panels that were intended to deflect the downwash in order to provide control and forward propulsion. The angle of each panel was controlled by levers mounted on either side of the pilot. Although Carnu, a French mechanic, achieved a historic first, the control of his machine was completely inadequate and it never developed into a practical helicopter.

July 22, 1908. Brequet built two variations of his first machine, which were hybrids: half-airplane and half-helicopter. Their side-by-side pair of large rotors were tilted toward the front to provide both lift and forward thrust. They actually flew on this date, going as high as 33 ft, but were wrecked on landing.

1910. Igor Sikorsky, in Kiev, Russia, constructed a 400-lb helicopter having two contrarotating, coaxial, canvas-covered rotors. The machine was barely able to lift its own weight. Realizing that there was much to be learned about helicopters, Sikorsky turned his attention to the fixed-wing airplane and in a relatively short time was successful in this endeavor.

July 1922. A helicopter designed by Emile Berliner, a German immigrant to the United States, rose to a height of 20 ft at College Park, Maryland. Although it flew forward for ~100 yd, the aircraft was marginally controllable. Two side-by-side rotors, which resembled large, rigid propellers, were mounted above a fuselage that had the appearance of a Fokker triplane. The helicopter had conventional horizontal and vertical tail surfaces with movable control surfaces, a winglike control surface between the wheels, and a third, smaller rotor to lift the tail.

1922. A helicopter designed by George deBothezat, a Russian immigrant to the United States, and built under support from the U.S. Army Air Service, was flown at McCook (later Wright) Field, Dayton, Ohio. More than 100 flight tests were performed, during which time a duration record of 2 min 45 sec was set and an altitude of 30 ft was achieved. This large machine, weighing 3600 lb and powered by a 180-hp Rhone engine, was lifted by four 26-ft-diameter rotors. As a result of its marginal performance and control, the Army discontinued its development.

January 9, 1923. The fourth autogyro designed by Juan de la Cierva, designated C.4, flew successfully at an airfield near Madrid, Spain. This was a significant breakthrough in the development of rotor design. The C.4 incorporated blades that, being hinged near the root, were free to flap up and down in response to the unsteady aerodynamic forces encountered in forward flight. This revolutionary approach to alleviating blade root stresses and unwanted moments was conceived by la Cierva after



FIGURE 2 First successful autogyro. Designed by Juan de la Cierva, the C.4 first flew on January 9, 1923, near Madrid, Spain. [Courtesy of The Smithsonian Institution.]

observing that one of his model autogyros with flexible blades was stable in flight, unlike his earlier full-scale machines. Other improvements followed the C.4. A rope-pull start was eliminated by deflecting the propeller slipstream with the horizontal tail up through the rotor. The start was further improved by the use of a clutch connected to the engine. Another major improvement was the addition of collective pitch control. This allowed the rotor to be spun up to a high speed in flat pitch. The pitch could then be suddenly increased, causing the autogyro to "jump" off of the ground using the energy stored in the rotor. La Cierva's articulated rotor system was the basis for all of the successful helicopter developments that followed. He was never to see it, however, as he met an untimely death at the age of 42 in the crash of a Dutch airliner as it was taking off from London's Croydon Aerodrome in December 1936. Figure 2 is a photograph of la Cierva's autogyro, which paved the way for the successful development of the helicopter.

October 1930. A helicopter designed by Corradino D'Ascanio of Italy took off near Rome and flew for one-half mile at an altitude of 59 ft for 8 min 45 sec. In so doing, it set the first records recognized by the Federation Aeronautique Internationale. Two coaxial, contrarotating rotors provided the lift. The blades, fragile in appearance, were controlled by flaps supported on booms trailing from each blade near the tip.

December 1935. At a military field outside Paris, Brequet demonstrated his latest helicopter to the Air Ministry. The craft had crashed almost 1 year earlier in a similar demonstration and had been rebuilt. Lifted by a pair of twin-bladed rotors, each having a diameter of 54 ft, it was flown to an altitude of 518 ft. Later it set an endurance record of 1 hr 2 min 50 sec and flew a distance of 27 miles. Built in collaboration with Giravions Durand, this Brequet-Durand helicopter won a million-

franc bonus from the Air Ministry but was far from being a practical helicopter.

June 26, 1936. On this date the FA-61 was flown for the first time. Designed by Henrich Focke, this German helicopter incorporated side-by-side rotors powered by a 160-hp engine. The shafts could tilt fore and aft for longitudinal control. The collective pitch of one rotor could be increased and the other decreased to provide roll control (differential collective pitch). After the first flight, the design continued to be improved, and on May 10, 1937, it demonstrated a perfect autorotational landing from 1130 ft with the engine off. Seven weeks later, the helicopter broke every official helicopter record in existence at the time. Many regard the FA-61 as the first truly successful helicopter. In 1938, a spectacular demonstration was performed when the female pilot Hanna Reitsch flew the FA-61 over a 3-week period before spectators in Berlin's large Deutschlandhalle sports arena. Figure 3 is a photograph of the FA-61.

1940. Another German helicopter, possibly more successful than the FA-61, the F1-282 designed by Anton Flettner, became operational with the German Navy. Named the Kolibri, this helicopter could fly at a speed of 90 mph and an altitude of 13,000 ft with a payload of 800 lb. Approximately 1000 of the F1-282 helicopters were produced. The configuration of the Kolibri was unique, incorporating side-by-side, closely spaced, intermeshing contrarotating rotors.

December 8, 1941. Igor Sikorsky personally flew the final configuration of his single rotor-tail rotor helicopter, the VS-300. Work on the machine began around 1938 and at one time it was flown with three tail rotors to provide control about all three axes. After encountering severe vibration and loss of control at speeds above ~30 mph, Sikorsky finally abandoned the two rotors mounted on outriggers at the tail in favor of the single vertical tail



FIGURE 3 FA-61 designed by Henrich Focke, 1936. [Courtesy of The Smithsonian Institution.]



FIGURE 4 Igor Sikorsky at the controls of his single rotor-tail rotor helicopter, the VS-300, December 1941. [Courtesy of The Smithsonian Institution.]

rotor, which is familiar today. The VS-300 employed a three-bladed, articulated rotor mounted above a fabric-covered, tubular-truss fuselage. [Figure 4](#) is a photograph of Sikorsky at the controls of the VS-300. Even before the problems with the VS-300 were solved, Sikorsky signed a contract with the U.S. Army to produce a much larger, operational helicopter. This machine, the R-4, was produced in quantity and provided valuable services for the U.S. military during the end of World War II.

Following Sikorsky's success, the helicopter developed rapidly with designs by notables such as Lawrence Young, which led to the formation of the Bell Helicopter Co., now a division of Textron, and the work of Frank Piasecki, who developed the first successful tandem-rotor helicopter, the HRP shown in [Fig. 5](#). Piasecki founded the Piasecki Helicopter Co., which was to become the Vertol Helicopter Co., now Boeing Helicopter Co. Surely the early pioneers who struggled simply to rise from the ground would marvel at the feats performed by the modern helicopter, epitomized by the first round-the-world flight of a helicopter



FIGURE 6 Modern, turbine-powered executive helicopter, the Sikorsky S-76.

accomplished in September 1982. The *Spirit of Texas*, a Bell Jet Ranger 206L-1, performed this feat, averaging almost 800 miles per day.

Today, the helicopter finds a variety of applications, from the support of offshore oil platforms to military combat to missions of mercy. The helicopter has saved countless lives, plucking persons from the rooftops of burning buildings, lifting sailors from sinking ships, rushing accident victims to a hospital, or rescuing flood victims from rising waters.

Typical of the modern helicopter are the four U.S. helicopters shown in [Figs. 6–9](#). The Sikorsky S-76 is an executive, turbine-powered helicopter designed to carry a crew of 2 and 12 passengers under instrument-flight conditions. This helicopter employs an articulated rotor, described later in this article. The McDonnell–Douglas (formerly Hughes) AH-64, the Apache, is typical of an attack helicopter. Equipped with a variety of armament, this helicopter utilizes a hingeless rotor and, being highly maneuverable, is designed to perform nap-of-the-earth operations under combat conditions. The Bell 206B is a light,



FIGURE 5 First successful tandem-rotor helicopter, the XHRS-X, with its designer Frank Piasecki at the controls.



FIGURE 7 Attack helicopter, the McDonnell–Douglas AH-64 Apache (formerly Hughes).



FIGURE 8 Modern, turbine-powered utility helicopter, the Bell 206B Jet Ranger.

utility type of helicopter that finds many applications. Fitted with a spray boom, it is used for defoliating power lines or for agricultural applications. Many police departments operate the 206B, as do hospitals for emergency use. This helicopter employs a teetering rotor, which is discussed later. The Boeing CH-47 Chinook is a large tandem, articulated-rotor helicopter designed to carry 33 troops at speeds of up to 158 knots. The civilian version of the Chinook, the Model 234, is used to support offshore oil platforms in the North Sea.

A future military helicopter is shown in Fig. 10. The RAH-66 Comanche, being developed jointly by Sikorsky and Boeing, is advertised as the world's most advanced helicopter and the cornerstone of the



FIGURE 10 The Sikorsky-Boeing RAH-66 Comanche.

U.S. Army's aviation modernization program. Powered by two turbine engines, it can cruise at 161 knots, dash to 172 knots, and fly sideways or backwards at 60 knots. It is designed to be stealthy, with a radar signature much smaller than the Apache (Fig. 7), which it will probably replace. The production version of the Comanche is flying and currently undergoing evaluation tests.

One of the latest civil helicopters being developed in the United States is shown in Fig. 11. The S-92 Helibus, built and currently undergoing certification by the Sikorsky Aircraft Corporation, is designed to serve a variety of commercial and utility needs. It has a capacity of up to 22 passengers and, in addition to serving as a general transport, it can support offshore oil operations or satisfy the needs of executive and aeromedical transport. The S-92, powered by two turbine engines, will have a cruising speed of approximately 155 knots, a hover ceiling, out of ground effect, of 6100 feet, and a range of approximately 480 nautical miles. The rotor is equipped with four blades and has a diameter of 56.25 ft. Its maximum takeoff gross weight is approximately 25,000 lb.



FIGURE 9 Large, tandem-rotor helicopter, the Boeing CH-47 Chinook.



FIGURE 11 The Sikorsky S-92 Airbus.

II. TECHNICAL CONSIDERATIONS

A. Rotor Velocity Field

As noted earlier, the success of the helicopter hinged on Juan de la Cierva's discovery of allowing the rotor blades to flap up and down (out of plane). To understand why this is necessary, refer to the top view of a rotor in forward flight, as depicted in Fig. 12.

Viewed from above, the rotor is pictured rotating counterclockwise, which is the U.S. convention. European rotors generally are designed to rotate in the opposite direction. The angular position of the blade ψ (azimuth angle) is measured from the downstream-most aft position. As the blade moves from $\psi = 0$, it advances into the flow, so that the linear velocity at any blade radius ωr and the forward velocity V of the helicopter add vectorially. Specifically at $\psi = 90^\circ$, as depicted in the figure, the velocity at the tip of the advancing blade is given by

$$V(90^\circ) = \omega R + V, \quad (1)$$

where ω is the rotational speed of the rotor in radians per unit of time and R is the tip radius. For ψ values between 180 and 360° , the rotor is retreating from the forward motion of the helicopter so that the velocity normal to a blade section is reduced by a component of V . Specifically, at $\psi = 270^\circ$, as shown in Fig. 12, the velocity at the tip of the retreating blade is

$$V(270^\circ) = \omega R - V. \quad (2)$$

Thus, everything else being equal, the advancing blade will experience a higher velocity, and hence higher lift, than that of the retreating blade. This imbalance in the

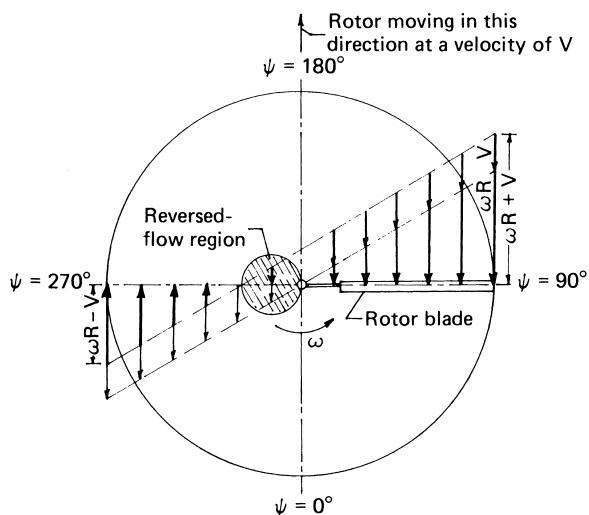


FIGURE 12 Top view of helicopter showing in-plane velocities affecting the rotor blades.

rotor forces will then produce a rolling moment on the helicopter, tending to roll it counterclockwise as viewed from the rear.

B. Reversed-Flow Region

For any ψ value and radial location r along the blade, the velocity normal to the leading edge of the blade is given generally by

$$V(\psi, r) = \omega r + V \sin \psi. \quad (3)$$

It is interesting that, for a small area of the disk on the retreating side, the above velocity is negative. The boundary of this reversed-flow region is found by letting the above velocity equal zero. It follows that

$$r/R = -\mu \sin \psi, \quad (4)$$

where μ is a parameter important for the analysis of helicopter aerodynamics. Known as the tip speed ratio, μ is defined as the ratio of the forward speed V to the rotor tip speed due to rotation ωR . The reversed-flow region is indicated in Fig. 12 for a μ of 0.3. In this region, the blade experiences a velocity into the trailing edge of the blade opposite to the flow direction for which an airfoil is designed to operate. The result can be a loss of blade lift in this region, leading to high vibratory loads. Generally, helicopters operate at tip speed ratios of less than 0.3, one of the reasons being the vibration limit imposed by the reversed-flow region.

C. Flapping

Now suppose that each blade is hinged at the hub so as to allow the blade to flap freely out of the plane of the rotor. This eliminates the rolling moment on the helicopter since a moment cannot be transferred through the hinge. The question then comes immediately to mind, however, of how the blade manages to remain more or less horizontal while sustaining a lift.

To answer this question, consider the blade at some arbitrary azimuth angle ψ , as shown in Fig. 13. The blade is pictured flapped up at some angle β . Generally, of course, the blade is moving at this instant, so there are inertia forces acting on the blade that contribute to the moment at the hinge. The sum of this moment and the moments caused by the aerodynamic lift, the blade weight, and the centrifugal force must equal zero at the hinge.

The flapping motion of the blade closely approximates first harmonic motion (in ψ) about a mean β angle β_0 , known as the coning. Writing the equation of motion for the blade about the flapping hinge for first harmonic flapping, it is found that the unsteady moment due to centrifugal force cancels the moment resulting from the spanwise distribution of inertia forces. It then follows that

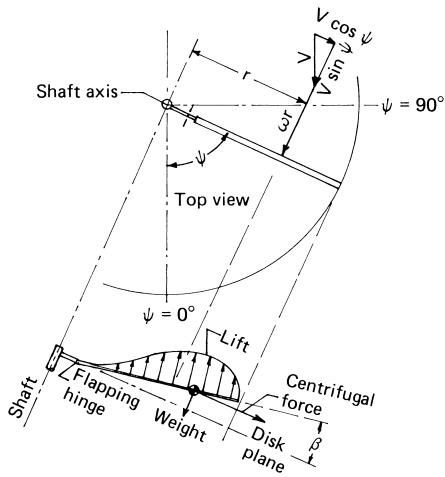


FIGURE 13 Forces and velocities affecting a flapping rotor blade.

the moment M_L about the flapping hinge caused by the aerodynamic lift on the blade is constant and equal to the sum of the weight moment M_W and the moment due to the centrifugal force with the blade coned at the angle β_0 . This can be expressed as

$$M_L = M_W + I_F \omega^2 \beta_0, \quad (5)$$

where I_F is the mass moment of inertia of the blade about the flapping hinge and β_0 is assumed to be a small angle. The weight moment is small in comparison with the moment resulting from the centrifugal force, so it is primarily the centrifugal force on the rotating blade that prevents it from folding upward.

In view of the foregoing, consider the following sequence of events starting with the rotor at rest. The blades will be drooped downward against droop stops with the rotor at rest. The pilot starts the engine and then engages a friction clutch to bring the rotor gradually up to speed. This is done with the collective pitch at a near-zero setting. In coming up to speed, the natural frequency of the helicopter, setting on its elastic landing gear, will at some point equal the one-per-revolution (1 per rev) frequency of inertia forces resulting from the rotor center of gravity being off the axis of rotation because of elastic deformation of the shaft and blades. This ground resonance was the downfall of many of the early experimenters and, even today, is carefully monitored in the initial testing of a new design.

When the rotor is up to speed, the pilot engages a jaw clutch to connect the engine and rotor positively. The pilot then increases the pitch collectively on the rotor, which develops lift on the blades. The blades begin to cone upward as a result of the lift. As the coning increases, the moment due to centrifugal force develops, which opposes the coning. An equilibrium coning angle is thereby reached where the centrifugal and blade weight moments balance the lift moment.

From Fig. 12, the lift on the blade varies with ψ , since the component of V normal to the blade depends on ψ . The lift also varies with ψ because of the flapping and control input. First, consider the effect of flapping. Figure 14 shows a section of the rotor blade under the influence of the linear velocity due to rotation ωr and the component of the helicopter's forward speed, normal to the leading edge of the blade, $V \sin \psi$. If the blade is flapping up, then, relative to the blade, an additional velocity vector downward must be added to $\omega r + V \sin \psi$. This results in a reduction $\Delta\alpha$ in the angle of attack of the blade section as shown. Hence, there is generated an aerodynamic moment opposing the flapping proportional to the flapping rate—a damping moment.

D. Main Rotor Control

With the unsteady centrifugal moment being proportional to the displacement and with the damping moment as described above, the motion of a flapping rotor blade closely resembles that of a second-order linear system with damping. It is well known that, if such a system is forced sinusoidally at its natural frequency, the displacement of the system will lag the force by 90° . Thus, to control the rotor, that is, to make the blade flapping peak up at some value of ψ , a force must be applied to the blades that peaks at $\psi = 90^\circ$. Such a control is achieved by cyclically varying the pitch of the blades using a swashplate mechanism, as shown schematically in Fig. 15.

The swashplate mechanism is composed of two plates. The one on the bottom is nonrotating, with the drive shaft passing through its center. This plate is gimbaled so that it can be tilted forward or back, or side to side, or raised or lowered by means of control rods. The upper plate is connected to and rotates with the shaft. It is also gimbaled and constrained mechanically so that its plane follows the plane of the lower plate.

Pitch-control arms attached to the blade are linked to the upper plate so that, as the attachment point on the

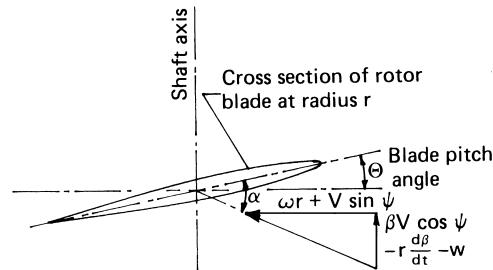


FIGURE 14 Influence of flapping on a section of the rotor blade, where α is the angle of attack, w induced velocity, and $\partial\beta/\partial t$ flapping rate.

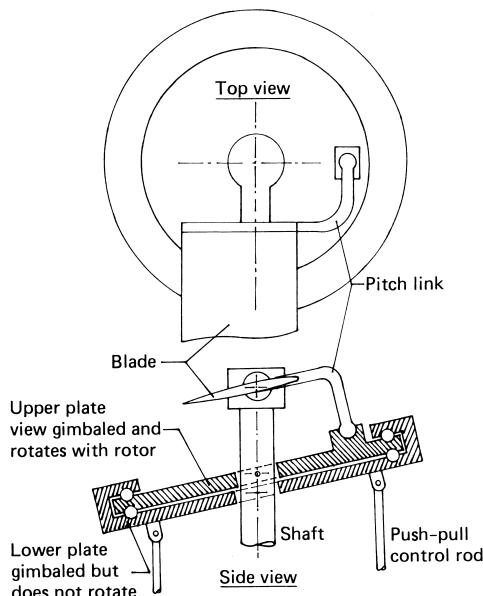


FIGURE 15 Swashplate.

upper plate moves vertically, following the lower plate, the pitch of the blade changes. Note that the attachment is advanced 90° ahead of the radialwise axis about which the blade pitches. (This pitch change is called feathering, and the axis about which it takes place is referred to as the feathering axis.)

Returning to the sequence of events described earlier, the pilot raises the collective pitch lever, which causes the swashplate to rise by a constant amount around the azimuth. The helicopter then rises vertically. To go forward the pilot moves the cyclic control stick forward. This causes the rear of the swashplate to move up and the nose to move down. Thus, as the blade comes to the right side ($\psi = 90^\circ$), its pitch angle will reach a minimum since the pitch arm is connected to the plate at the nose ($\psi = 180^\circ$). Thus, because of the cyclic pitch, the aerodynamic lift on the blade will vary sinusoidally around the azimuth, experiencing a maximum value at $\psi = 270^\circ$ and a minimum value at $\psi = 90^\circ$. As explained earlier, the blade displacement (flapping) lags the force by 90°. Therefore, it flaps down in the front and up in the back so that the plane prescribed by the path of the blade tips, the tip-path plane, tends to follow the plane of the swashplate.

The rotor thrust vector is closely normal to the tip-path plane. Thus, in the preceding example, as the tip-path plane noses downward, the rotor thrust vector tilts forward, causing the helicopter to accelerate in that direction.

E. Tail Rotor Control

Yaw control, that is, the control to nose the helicopter fuselage left or right, is accomplished by means of the

tail rotor (for a single rotor-tail rotor machine). A major function of the tail is, first, to counteract the torque of the main rotor. Referring once again to Fig. 12, with a counterclockwise main rotor rotation, the tail rotor thrust must be to the right and of a magnitude such that the product of its thrust T_T and the distance l_T from its axis to the main rotor axis is equal to the main rotor torque Q . Knowing the power P required by the main rotor, we readily obtain the torque from

$$P = \omega Q. \quad (6)$$

This required tail rotor thrust can diminish in forward flight if a vertical fin is present at the tail. In this case, if the helicopter yaws, a transverse aerodynamic force is generated on the fin, producing a moment that opposes the yaw.

The tail rotor has a collective pitch control, which is connected to rudder pedals at the pilot's feet. Pushing forward on the right pedal will cause the collective pitch of the tail to decrease. The tail rotor thrust will therefore decrease and the torque of the main rotor, now unbalanced, will cause the helicopter to nose, or yaw, to the right.

The conventional tail rotor resembles a propeller mounted near the rear and on the side of the tail cone (Fig. 16). The tail cone is that part of the fuselage extending back from the cabin area. Currently, the tail rotor is the most common means used to counteract the torque of the main rotor and was the configuration employed by Igor Sikorsky in his first successful helicopter. The open tail rotor spinning at a high speed is dangerous when operating on the ground, and alternative means to counteract the torque have been tried over the years. The tandem rotor configuration does not require a tail rotor because the two main rotors rotate in opposite directions so that the net torque on the fuselage is zero.

In recent times, two other devices to counteract the torque of a single-rotor helicopter have been developed. The first, and most proven, is the fan-in-tail configuration, or Fenestron, as named by the Aerospatiale Company of France and employed on a number of its helicopters.

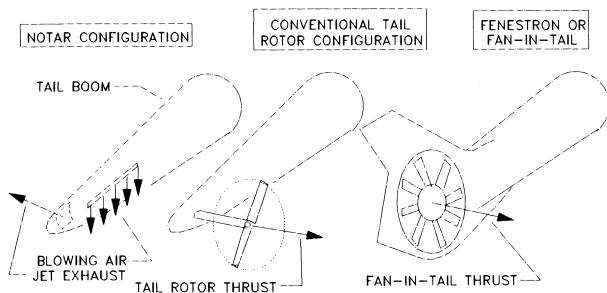


FIGURE 16 Schematic drawings of various tail thrusters mounted on the end of single-rotor tail booms.

Here, a ducted fan is placed in the tail cone as shown in Fig. 16. As in the case of the conventional tail rotor, the thrust provided by the ducted fan is controlled by varying the pitch of the blades. Although the power required by this configuration is normally higher than the conventional tail rotor, advocates of the fan-in-tail point to the greatly increased safety afforded by the configuration.

The most recent antitorque configuration has been developed by the McDonnell-Douglas Helicopter Co. and has been given the name NOTAR, an acronym standing for *no tail rotor*. A sketch of NOTAR is also presented in Fig. 16. Referring to this figure, a fan, mounted in the tail boom, compresses the air, forcing some of it out of a long slot along the right side of the boom. The slot directs the thin sheet of air downward and tangential to the rounded surface of the boom. Because of the coanda effect, the sheet follows around the boom, causing a circulatory flow around the boom. The interaction of this circulation with the downwash from the main rotor produces a lift on the boom directed to the right, according to the Kutta-Joukowski theorem. The remainder of the air compressed by the fan exits to the left through a nozzle at the very rear of the boom. This jet of air is throttled by a sleeve which rotates to partially close the nozzle. The combination of the boom circulatory lift and the nozzle thrust provides the antitorque while the throttling of the nozzle provides yaw control.

F. Tandem Rotor Control

Each rotor of a tandem-rotor helicopter, such as the Boeing CH 47 Chinook (Fig. 9) is equipped with a swashplate mechanism so that the tip-path plane of each rotor can be separately controlled. Yaw control is accomplished by tilting the thrust vector of the front rotor to one side and the thrust vector of the rear rotor to the opposite side. To nose the helicopter up or down (i.e., pitch control), a combination of cyclic control and differential collective pitch is used. For example, to move forward, the tip-path planes of both rotors are tilted forward. Also, however, the collective pitch on the aft rotor is increased and the collective pitch on the forward rotor decreased. This reduces the thrust on the forward rotor while increasing the thrust on the aft rotor, which results in a nose-down pitching moment.

G. Helicopter Performance

The power required for level flight by a helicopter is composed mainly of three components: the parasite power, induced power, and profile power. In addition to these main components, increments are required to cover transmission losses, cooling, and the tail rotor.

The parasite power P_{par} , the power required to overcome the parasite drag of the helicopter in forward flight, is given simply by

$$P_{\text{par}} = DV, \quad (7)$$

where D is the drag of the helicopter at a forward speed of V . In hover, the parasite power obviously vanishes.

The profile power is the power required to overcome the drag of the rotor blades opposing the rotation of the rotor. The power of a helicopter is frequently expressed in terms of a dimensionless coefficient C_p , defined by

$$C_p = \frac{P}{\rho AV_T^3}, \quad (8)$$

where ρ is the mass density of the air, A the rotor disk area πR^2 , and V_T the rotor tip speed due to rotation ωR . The profile power coefficient $C_{p,p}$ can be calculated approximately from

$$C_{p,p} = \frac{\sigma C_D}{8}, \quad (9)$$

where σ , the rotor solidity, is the ratio of blade area to disk area; C_D is an average blade section drag coefficient having a typical value of 0.01.

The induced power P_i represents the penalty paid to produce lift and is given by the product of the rotor thrust and the velocity w induced normal to the rotor disk.

From momentum principles, W for the hovering case can be obtained from

$$w = \sqrt{\frac{T}{2\rho A}}, \quad (10)$$

where T is the rotor thrust, which is nearly equal to the helicopter weight. In forward flight above ~ 30 mph, w is found from

$$w = \frac{T}{2\rho AV}. \quad (11)$$

Figure 17 presents a breakdown of the required power as a function of forward speed for a typical helicopter. Note that the profile power is nearly constant with speed while the induced power is a maximum in hover, decreasing rapidly as the forward speed increases. The parasite power is zero at hover but becomes the predominant contribution to the total power at high forward speeds.

Knowing the power required by a helicopter to maintain level flight and the installed engine power, one can determine various performance items, such as rate of climb and hover ceiling and range. Table I lists physical characteristics and performance for several U.S. and European helicopters of different sizes.

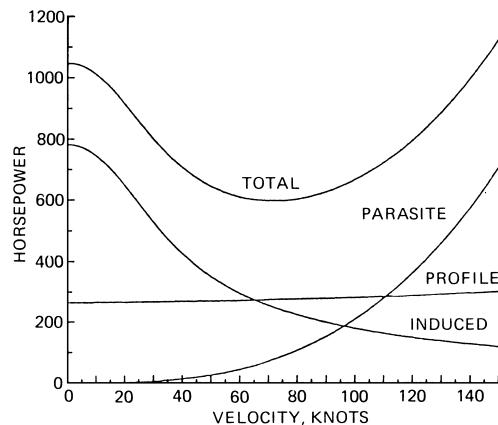


FIGURE 17 Power components required by a typical attack helicopter (gross weight, 10,000 lb) in level forward flight.

H. Design Variables

The square–cube law states that any area of an aircraft varies as the square of a characteristic length while the volume and weight vary as the cube. These relationships are, of course, only approximate, and for other reasons a particular aircraft may be more densely packed than another of the same size, or advances in material technology can change the relationship between volume and weight. Nevertheless, the square–cube law provides a basis for correlating design trends of helicopters (and fixed-wing airplanes) with gross weight. Assuming the rotor disk area to be proportional to a characteristic length L squared, and the weight W proportional to L cubed, it follows that the disk loading is given by

$$\frac{T}{A} \propto W^{1/3}. \quad (12)$$

Figure 18 presents the disk loading for several modern helicopters as a function of the gross weight, and indeed, as shown, a curve that takes T/A proportional to the cube root of the weight fits the data points fairly well. Note that small helicopters have disk loadings as low as ~ 3 pounds per square foot (psf), whereas the disk loadings of large helicopters are as high as 15 psf.

As a result of the square–cube law and the induced power relationship given earlier, the power loading (pounds per horsepower) decreases as the helicopter size, and hence disk loading, increases. One would estimate the power loading to vary inversely with the square root of the disk loading. This trend is verified approximately by the data and the curve presented in Fig. 19.

The lift of a wing can be expressed in terms of a dimensionless lift coefficient C_L given by

$$C_L = \frac{L}{\frac{1}{2}\rho V^2 S}, \quad (13)$$

where L is the lift and S the planform area of the wing (projected area looking down on the wing). For a given wing, C_L is a function only of the angle of attack of the wing, the angle between the flow velocity and the plane of the wing. Thus, knowing C_L , one can calculate wing lift for any size of wing, airspeed, and altitude (air density). A wing can reach only a maximum value of $C_{L\max}$. For angles of attack higher than that corresponding to $C_{L\max}$, the airflow separates from the upper surface of the wing and the wing is said to be stalled. Similarly, the lift of

TABLE I Helicopter Characteristics

Characteristic	Aerospatiale AS350L Ecureuil	Aerospatiale SA365F Dauphin 2	Augusta SH-3D/H Sea King	Bell UH-1H Huey	Bell AH-1S Cobra	Boeing Vertol CH-47C Chinook	Hughes AH-64A Apache	Messerschmitt–Bolkow–Blohm BK 117	Sikorsky CH-53E Super Stallion
Gross weight (lb)	4,299	8,600	20,500	9,500	10,000	50,000	20,500	6,283	73,500
Maximum continuous hp	592	627	1,250	1,250	1,300	3,000	1,536	410	3,696
Main rotor diameter (ft)	35.07	39.13	62.0	48.0	44.0	60.0(2)	48.0	36.08	79.0
HOGE (ft) ^a	7,380	5,740	8,200	6,004	3,750	5,900	12,400	8,200	9,500
Service ceiling (ft)	15,600	15,000	14,700	12,700	12,200	8,500	20,500	14,760	18,500
R/C (fpm) ^b	1,555	1,515	2,200	1,700	1,625	1,485	2,880	1,950	2,750
V _{NE} (knots) ^c	147	165	144	128	170	160	197	150	170
Maximum cruise (knots)	125	140	118	110	128	158	145	136	150

^a Hover ceiling out of ground effects.

^b Rate of climb in feet per minute.

^c Velocity to never exceed.

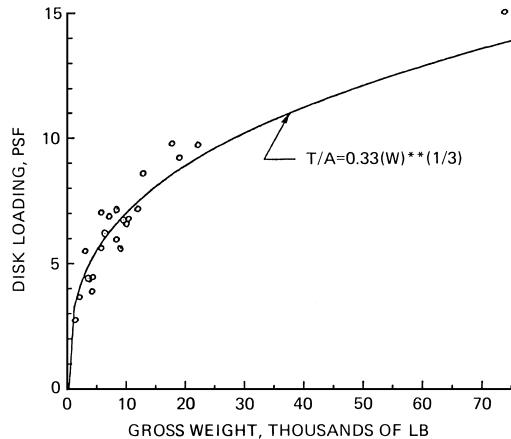


FIGURE 18 Current helicopter disk loadings as a function of gross weight.

a helicopter rotor blade can be expressed in terms of an average C_L . One would not expect the dimensionless coefficient to vary with helicopter size and, indeed, the data for existing helicopters show no trend of C_L with gross weight. Most helicopters are designed to operate at average rotor lift coefficients of between 0.4 and 0.55. One can calculate C_L approximately from

$$C_L = \frac{6C_T}{T}, \quad (14)$$

where C_T is the rotor thrust coefficient, defined by

$$C_T = \frac{T}{\rho AV_T^2}. \quad (15)$$

Since $C_{L\max}$ has a typical value of 1.5, the average rotor C_L of 0.55 or less allows some margin for control. If the rotor were designed to operate close to $C_{L\max}$, increasing the collective pitch to increase the rotor thrust might cause the rotor to stall, resulting, instead, in a decrease of thrust.

Average lift coefficients for tail rotors required to balance the torque of the maximum continuous power tend to be somewhat higher than those used for the main rotor. Typically, these vary from ~ 0.5 up to ~ 0.7 . Disk loadings for tail rotors are considerably higher than those for the main rotor and range from ~ 10 psf for gross weights of ~ 4000 lb up to 25 psf for large helicopters.

I. Types of Main Rotors

Three types of main rotors are popular with helicopter manufacturers: (1) articulated, (2) hingeless, and (3) teetering. For example, most Bell helicopters employ teetering rotors. Boeing and Sikorsky use articulated rotors, and Messerschmitt–Boelkow–Blohm uses a hingeless rotor.

The articulated and hingeless rotors are aerodynamically similar. Both types are designed to allow the blades to flap, not only out of plane but also in plane to relieve in-plane Coriolis forces, which result from the out-of-plane flapping. The in-plane motion is referred to as lead-lag motion. It is an unsteady angular movement of the blade that is superimposed on the steady rotor rotation.

The articulated rotor has flapping and lead-lag hinges to allow for the blade motion. The hingeless rotor employs composite blades, “flex” straps, or elastometric bearings to allow for the blade motion. Most manufacturers have developed hingeless designs, and this type of rotor will probably find increasing use in the future.

The teetering rotor is a two-bladed rotor that, like a seesaw, pivots about the top of the rotor shaft so that as one blade moves up the other moves down.

There are advantages and disadvantages for each system. The teetering rotor system is mechanically relatively simple and easy to maintain. Its aerodynamic drag is also relatively low. However, roll and pitch control for this system is accomplished by tilting the thrust vector so as to provide rolling or pitching moments about the helicopter's center of gravity. These control moments vanish as the rotor thrust approaches zero. Thus, for applications in which low- g maneuvers are required, the use of a teetering rotor is questionable. As the rotor thrust decreases, the tail rotor, being above the center of gravity, can cause the helicopter to roll. Instinctively, the pilot may counter with lateral cycle control. This will cause the rotor to tilt to the side, but without thrust the rolling of the helicopter is not stopped. Under extreme conditions, the rotor can flap excessively relative to the fuselage, causing the rotor hub to strike and sever the shaft. Such a catastrophic encounter has become known as mast bumping. This is not to imply that the teetering rotor is inherently unsafe, but the pilot must be aware of this characteristic and be trained in proper

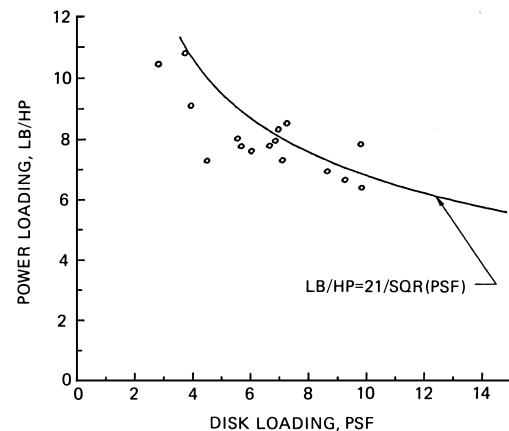


FIGURE 19 Effect of disk loading on power loading.

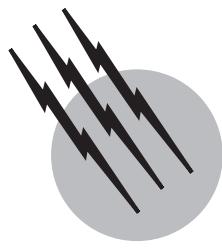
recovery techniques. Indeed, as mentioned earlier in this article, the *Spirit of Texas*, which made the round-the-world flight, was a Bell 206-L-1 incorporating a teetering rotor.

SEE ALSO THE FOLLOWING ARTICLES

AIRCRAFT AVIONICS • AIRCRAFT INSTRUMENTS •
AIRCRAFT PERFORMANCE AND DESIGN • AIRCRAFT
SPEED AND ALTITUDE • AIRPLANES, LIGHT • FLIGHT
(AERODYNAMICS) • RAMJETS AND SCRAMJETS • V/STOL
AIRPLANES

BIBLIOGRAPHY

- Boyne, W. J., and Lopez, D. S., eds. (1984). "The Age of the Helicopter—Vertical Flight," Smithsonian Institution Press, Washington, DC.
- Bramwell, A. R. S. (1976). "Helicopter Dynamics," Arnold, London.
- Gessow, A., and Myers, G. C., Jr. (1967). "Aerodynamics of the Helicopter," Ungar, New York (originally published 1952).
- Johnson, W. (1980). "Helicopter Theory," Princeton University Press, Princeton, NJ.
- McCormick, B. W. (1967). "Aerodynamics of V/STOL Flight," Academic Press, New York.
- Prouty, R. W. (1986). "Helicopter Performance Stability and Control," PWS Engineering, Boston.
- Young, W. (1982). "The Helicopters," Epic of Flight Series, Time-Life Books, Alexandria, VA.



Structural Analysis, Aerospace

David H. Allen

Texas A&M University

- I. Historical Introduction
- II. The Role of Structural Analysis
- III. Mechanics of Aerospace Structures
- IV. Analysis of Advanced Beams
- V. Advanced Analysis Techniques

GLOSSARY

- Advanced beam** Long, slender structural member, possibly of nonhomogeneous material makeup, subjected to multiaxial bending, extension, and/or thermal loads.
- Aerospace structure** Structure whose usefulness diminishes significantly with increasing weight.
- Composite** Structural component composed of two or more materials bonded together in some fabrication process, thus resulting in a composite material with a high strength-to-weight ratio.
- Constitution** That set of constraints that distinguish a body from any other body of identical geometric shape; also called material properties.
- Continuum mechanics** Approach to the analysis of media that ignores the microstructural features such as molecular bonds and seeks to model the bodies as continuous.
- Elastic** Constitutive term meaning that the stress and strain are uniquely related at a material point.
- Finite element method** Most powerful structural analysis technique currently in use; obtained by discretizing the actual structure into subcomponents (called finite elements) for purposes of analysis.

Statically determinate Term used when the internal loads can be determined at all points in a body by using Newton's laws of motion and without recourse to the material makeup of the body.

Statically indeterminate Converse of statically determinate.

Structural analysis Step in the structural design process whereby predictions of the temperature, deformation, and internal loads of some structure of predetermined shape are obtained.

Structural design Process whereby a structure is iteratively modified until all design constraints are satisfied.

Variational methods Class of mathematical techniques that can be utilized for performing structural analysis.

AEROSPACE structures include atmospheric flight vehicles, such as shown in Fig. 1, orbital vehicles, such as shown in Fig. 2, and interplanetary probes, such as Explorer I. Analysis of this class of structures consists of the theoretical prediction of deformations, stresses, and temperatures throughout the vehicle, given the external thermomechanical loading conditions and the geometric shape of the structure. The object of this analysis is to

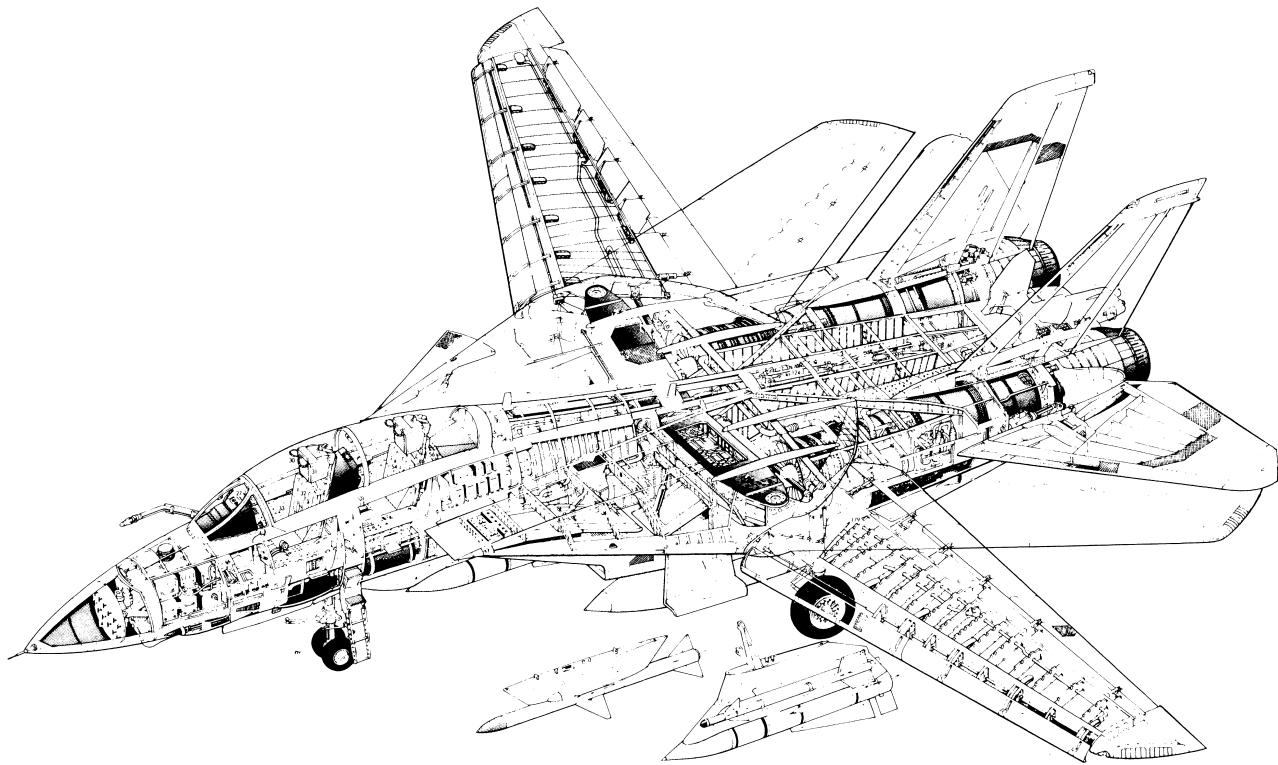


FIGURE 1 Grumman F-14. [Courtesy of the U.S. Navy.]

predict analytically such catastrophic events as excessive deformations, buckling, and fracture, which can lead to structural failure. Because it is extremely important to minimize mass in these structures, relatively small margins of safety are allowed between analytical predictions and the criteria for structural failure. Therefore, it is necessary that the theoretical tools used in the analysis result in accurate predictions. This requirement often necessitates the development of highly complex algorithms requiring numerous well-trained structural analysts and a large capital outlay compared to the actual fabrication cost of the structure.

I. HISTORICAL INTRODUCTION

The first manned flight of a motorized vehicle by the Wright Brothers at Kitty Hawk, North Carolina, on December 17, 1903, is probably the most technologically significant achievement in history. It marked the transition from an age of travel on foot, and by horse and wagon, to Neil Armstrong's footprints on the Moon little more than 65 years later. However, the concepts embodied in the design and analysis of flight structures date back hundreds and in some cases even thousands of years.

The analysis of aerospace structures is rooted in the fundamental axioms of continuum mechanics. The first

exposition of one of these axioms was published in 1676 by Robert Hooke (1635–1703). He stated that the power of any spring is in the same proportion with the tension thereof, thus stating the first constitutive equation for linear elastic bodies.

A short time later, Sir Isaac Newton (1642–1727), probably the greatest scientist of all time, reported his axioms governing conservation of momentum. These were expounded in his text “Philosophiae Naturalis Principia Mathematica,” published in 1687. The elastic field problem was finally reported by Navier (1785–1836) in 1821 and by Cauchy (1789–1857) in 1822. At the same time that Navier and Cauchy were investigating the elastic field problem, Fourier (1768–1830) was constructing the equations modeling the temperature field in a body of general shape. His results on heat conduction were reported in his classical treatise, “La Theorie Analytique de la Chaleur,” in 1822. The remainder of the nineteenth century was marked by numerous achievements in elasticity theory by such great scientists as Poisson (1781–1840), Navier (1785–1836), Green (1793–1841), Clapeyron (1794–1864), Saint-Venant (1797–1886), Duhamel (1797–1872), Airy (1801–1892), Helmholtz (1821–1894), Kelvin (1824–1907), Maxwell (1831–1879), Mohr (1835–1918), and Rayleigh (1842–1919).

The twentieth century is marked by numerous milestones in aerospace technology such as the first liquid-fuel

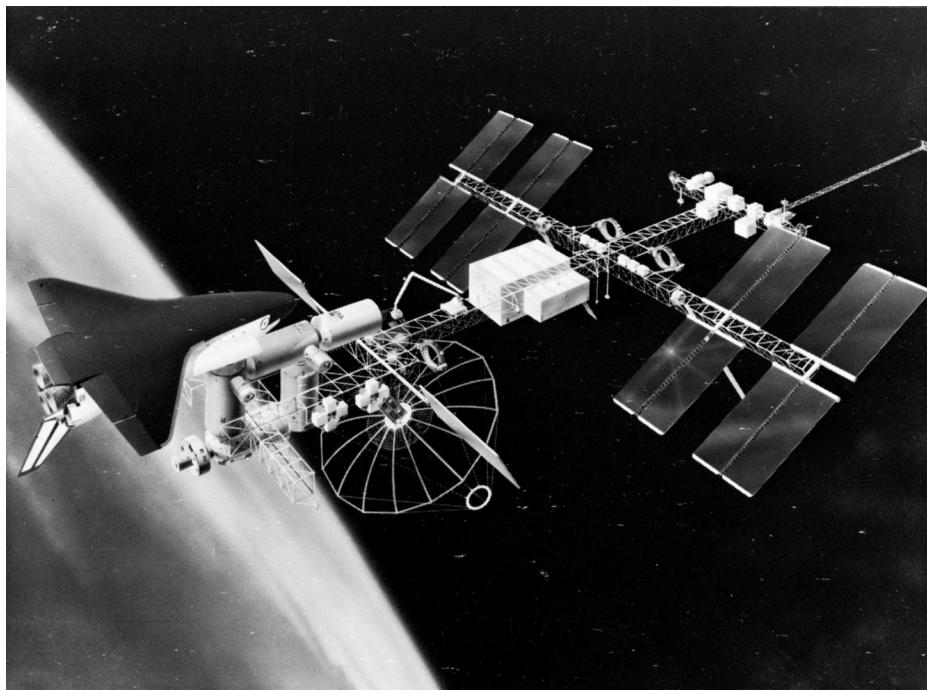


FIGURE 2 Artist's concept of orbiting space station. [Courtesy of the National Aeronautics and Space Administration.]

rocket flight (1926), the first jet engine-powered flight (1939), the first supersonic flight (1947), the first artificial satellite launch (1957), the first manned flight (1961), and the first space shuttle flight (1981). All of these achievements posed many engineering problems and challenges, but for the structural analyst the foremost of these was to achieve minimum weight with maximum safety. Many of the advances have come with the development of advanced materials such as fibrous composites, which have very high strength-to-weight ratios.

Improvements in analytical techniques were basically driven by the advent of the high-speed digital computer, thus making the stiffness method, which has now evolved into the finite element method, more computationally efficient and usable in everyday structural analysis.

II. THE ROLE OF STRUCTURAL ANALYSIS

Structural design is the process whereby the structural engineer begins with little more than a set of loads and design constraints and proceeds iteratively to obtain a structural configuration that satisfies all of these constraints. As shown in Fig. 3, structural analysis is one step in the structural design process. Structural analysts are usually given a well-defined structural geometry. It is their task then to determine whether the current configuration satisfies the

design constraints. These design constraints may include such things as fracture criteria, excessive deformations, size and weight constraints, and even cost constraints. If the analysis results in predicted failure of the structure due to any of the design constraints, then the design geometry must be modified. This process is then carried out iteratively until all of the design constraints are satisfied.

As indicated in Fig. 3, it is possible to check the design experimentally rather than analytically. However, for

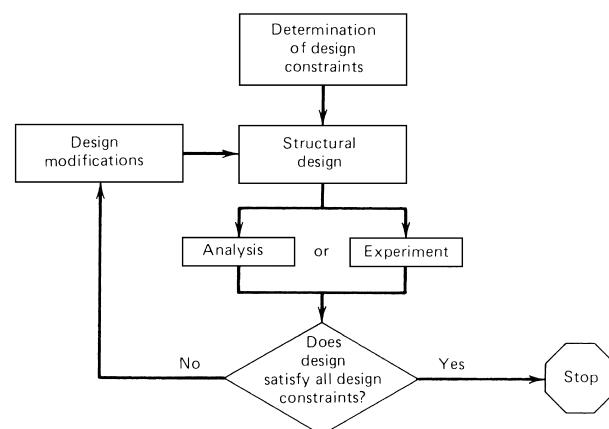


FIGURE 3 The structural design process. [From Allen, D. H., and Haisler, W. E. (1985). "Introduction to Aerospace Structural Analysis," Wiley, New York.]

aerospace structures this is often too costly. Therefore, analysis is often a necessary step in the structural design procedure.

Because one can never completely nullify all error sources, it is customary to design structures to withstand some ultimate loading configuration more stringent than the actually expected allowable load. Thus, a factor of safety SF is defined by

$$SF = \text{ultimate failure load/allowable load}, \quad (1)$$

where the ultimate failure load is that load that will cause the structure to fail to satisfy any of the design constraints. Another term that is often used is the margin of safety MS, defined by

$$MS = SF - 1. \quad (2)$$

The value of the safety factor used in a particular analysis will depend on many factors, such as the accuracy of the theory involved, the range of variability of material properties, and the cost of materials. In aerospace structures, where, because of mass production, engineering analysis may be budgeted at hundreds of times the cost of a single aircraft, relatively sophisticated analytic techniques are within cost constraints, and safety factors may be as low as from 1.25 to 1.5 for manned vehicles and from 1.1 to 1.25 for unmanned vehicles.

Although the basic design process described in Fig. 3 applies in the aerospace industry, the procedure is quite involved because of the complexity of the structures produced. There are usually several large groups of engineers, often numbering in the hundreds, who work in several separate groups. These include the design, aerodynamics, aeroelasticity, materials, and weights groups, as well as the structures group. The design process is generally interactive in nature and can be complicated by this interaction. For example, in a typical design process the design group might conceive of a horizontal tail unit for a fighter aircraft. This design would then be forwarded to the aerodynamics group to determine the external lifting and drag loads on the structure. Given these loads, the materials and structures group would then produce an internal structure capable of withstanding the loads. The weights group would then determine the weight of the structure as designed. If, for example, the weight is found to exceed the originally conceived weight designed for, it would then be necessary to modify the design. The decision as to how this modification should be made might involve one or all of the groups. The materials group might propose a lightweight composite material, but if this exceeds the cost constraint, even further design modification might be necessary. A complete design of the external geometry would certainly affect all groups and might lead to a new configuration that exceeds some other design con-

straint. Therefore, in complex aerospace structures there must be a great deal of interaction between the various engineering groups.

III. MECHANICS OF AEROSPACE STRUCTURES

Aerospace structures are analyzed by using field equations obtained from continuum mechanics and thermodynamics. These equations are algebraic or partial differential equations that apply at all points x_j and for all times t in a continuous body with interior V and surface S , as shown in Fig. 4. (Lowercase subscripts imply three components ranging from 1 to 3, thus representing Euclidean space.) The equations are in general (1) conservation of momentum (kinetics), (2) strain-displacement relations (kinematics), (3) stress-strain-temperature relations (constitution), and (4) conservation of energy (thermodynamics).

The difficulty of the analytic scheme will depend in large part on the level of complexity at which the preceding equations must be applied in order to obtain accurate theoretical models. For example, the shuttle may be considered to be a rigid body for purposes of calculating launch performance characteristics, thus simplifying this portion of the analysis. However, this assumption will give no insight about whether structural components within the shuttle can be expected to fail during launch.

For most aerospace vehicles it is sufficiently accurate to assume that the motions are infinitesimal and the constitution is linear elastic. These assumptions will in most cases produce a linear field problem when items 1 through 4 are cast in a mathematical context. Furthermore, it is usually acceptable to uncouple the conservation of energy from the remaining equations by assuming that the body of interest is rigid for purposes of calculating temperature. In the case where boundary conditions are radiative (as is the case in space), the heat conduction problem resulting from

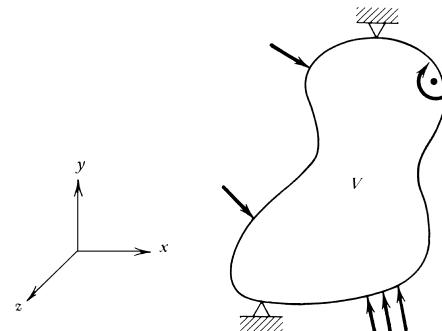


FIGURE 4 General body subjected to complex loading. [From Allen, D. H., and Haisler, W. E. (1985). "Introduction to Aerospace Structural Analysis," Wiley, New York.]

the conservation of energy will be the only nonlinearity in the analysis. Otherwise, for convective boundary conditions (as is usually the case in atmospheric conditions), the entire thermomechanical problem is linear and has a unique solution.

A. Field Variables

The quantities to be derived from the analysis are called field variables, and the analysis will result in a prediction of these quantities at all points x, y, z and for all times t of interest in V . The field variables of interest are stress, strain, displacement, and temperature. Once the analysis is complete these variables may be substituted into any constraint equations (such as failure criteria, mass restrictions, and deformation constraints) to ensure that the proposed structural configuration satisfies all design requirements.

1. Stress Tensor

Consider a point 0 in the interior of a body V . Newton's third law requires that the internal reactions at point 0 must be replaced with external force \mathbf{F} if a cutting plane A is passed through point 0, as shown in Fig. 5. The concept of stress is now introduced. The normal stress on plane A and at point 0 is defined by

$$\sigma_n \equiv \lim_{\Delta A \rightarrow 0} (\Delta F_n / \Delta A) = dF_n / dA, \quad (3)$$

where the symbol \equiv means defined to be.

The shear stress is defined by

$$\sigma_s \equiv \lim_{\Delta A \rightarrow 0} (\Delta F_s / \Delta A) = dF_s / dA. \quad (4)$$

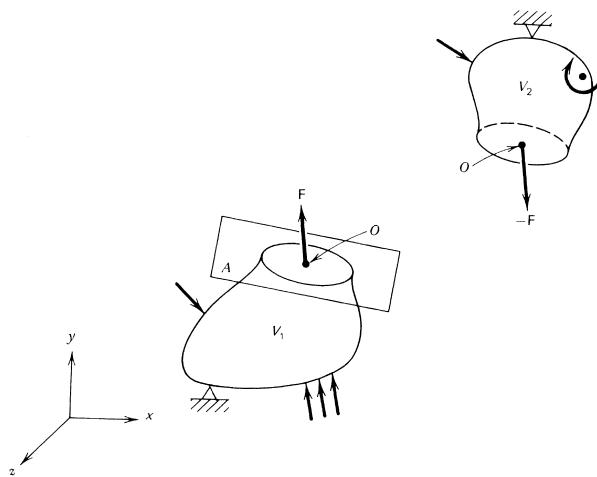


FIGURE 5 Free-body diagram showing internal forces. [From Allen, D. H., and Haisler, W. E. (1985). "Introduction to Aerospace Structural Analysis," Wiley, New York.]

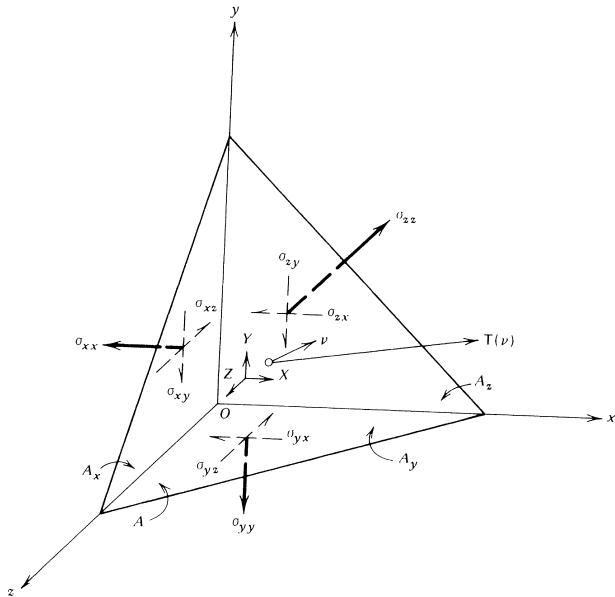


FIGURE 6 Description of the stress tensor at material point 0. [From Allen, D. H., and Haisler, W. E. (1985). "Introduction to Aerospace Structural Analysis," Wiley, New York.]

It can be shown that if three mutually perpendicular cutting planes are passed through point 0, the resulting nine components of stress, $\sigma_{xx}, \sigma_{yy}, \sigma_{zz}, \sigma_{yz}, \sigma_{zy}, \sigma_{xz}, \sigma_{zx}, \sigma_{xy}$, and σ_{yx} , as shown in Fig. 6 are sufficient to determine the components of the stress tensor on any other plane A at point 0. Therefore, there are nine unique components of stress to be determined at each point in the body V . The stress tensor represents a measure of the load intensity, or pressure, at each point in a body.

2. Strain Tensor

Consider again the point 0 in body V , as shown in Fig. 7. For infinitesimal deformations, the strain tensor is defined by

$$\begin{aligned} \varepsilon_{xx} &\equiv \frac{\partial u}{\partial x}, & \varepsilon_{yy} &\equiv \frac{\partial v}{\partial y}, & \varepsilon_{zz} &\equiv \frac{\partial w}{\partial z}, \\ \varepsilon_{yz} &\equiv \frac{1}{2} \left(\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y} \right), & \varepsilon_{xz} &\equiv \frac{1}{2} \left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x} \right), \\ \varepsilon_{xy} &\equiv \frac{1}{2} \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right), \end{aligned} \quad (5)$$

where u , v , and w are components of the displacement vector. Thus, there are six unique components of strain at each point in the body V . The strain tensor represents a measure of the deformation intensity, or deformation per unit length, at each point in a body.

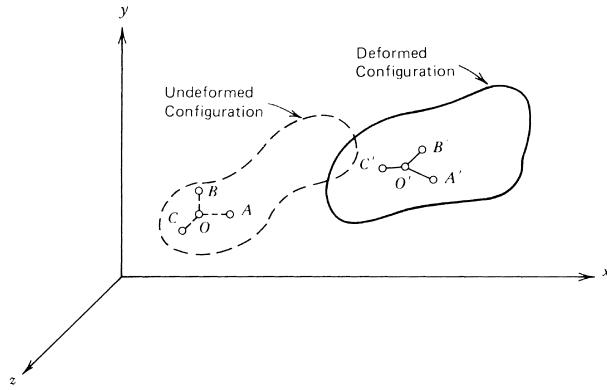


FIGURE 7 Three-dimensional body in undeformed and deformed configurations. [From Allen, D. H., and Haisler, W. E. (1985). “Introduction to Aerospace Structural Analysis,” Wiley, New York.]

B. Conservation Laws

For nonablating bodies undergoing infinitesimal deformations, it can be shown that conservation of mass is identically satisfied. Therefore, the nontrivial conservation laws in most aerospace structural applications are conservation of momentum and conservation of energy.

1. Conservation of Momentum

Conservation of linear momentum will lead to the following equations at every point in V :

$$\begin{aligned} \frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{yx}}{\partial y} + \frac{\partial \sigma_{zx}}{\partial z} + X &= \rho \ddot{u}, \\ \frac{\partial \sigma_{xy}}{\partial x} + \frac{\partial \sigma_{yy}}{\partial y} + \frac{\partial \sigma_{zy}}{\partial z} + Y &= \rho \ddot{v}, \\ \frac{\partial \sigma_{xz}}{\partial x} + \frac{\partial \sigma_{yz}}{\partial y} + \frac{\partial \sigma_{zz}}{\partial z} + Z &= \rho \ddot{w}, \end{aligned} \quad (6)$$

where X , Y , and Z are the components of the body force vector per unit volume (such as gravitational loads) and ρ is the mass density. For quasi-static conditions, the inertial term on the right-hand side of (6) may be neglected.

Furthermore, conservation of angular momentum on the surface S will yield

$$\begin{aligned} T_x &= \sigma_{xx} v_x + \sigma_{yx} v_y + \sigma_{zx} v_z, \\ T_y &= \sigma_{xy} v_x + \sigma_{yy} v_y + \sigma_{zy} v_z, \\ T_z &= \sigma_{xz} v_x + \sigma_{yz} v_y + \sigma_{zz} v_z, \end{aligned} \quad (7)$$

where T_x , T_y , and T_z are the components of the traction vector (force per unit area) and v_j the components of a unit outer normal at every point on S .

Finally, conservation of angular momentum will result in the following at every point in $V + S$:

$$\sigma_{yz} = \sigma_{zy}, \quad \sigma_{xz} = \sigma_{zx}, \quad \sigma_{xy} = \sigma_{yx}. \quad (8)$$

(This equation assumes that body moments may be neglected.) Thus, there are only six unique components of the stress tensor, which is said to be symmetric.

2. Conservation of Energy

For most realistic applications involving elastic materials (sometimes excluding dynamic circumstances), conservation of energy will lead to the Fourier heat conduction equation:

$$\frac{\partial}{\partial x} \left(k \frac{\partial T}{\partial x} \right) + \frac{\partial}{\partial y} \left(k \frac{\partial T}{\partial y} \right) + \frac{\partial}{\partial z} \left(k \frac{\partial T}{\partial z} \right) = \rho C_v \dot{T} - r, \quad (9)$$

where k is the thermal conductivity, C_v the specific heat at constant volume, and r the internal heat source per unit volume.

C. Constitution

The description of the field problem is completed with the specification of constitutive equations. These equations reflect the material properties of the various components utilized throughout the structure. In general, these are relations among stress, strain, and temperature that apply at each material point in the structure. For example, if a typical structural component is subjected to a prescribed uniaxial deformation history, as shown in Fig. 8, a cross-plot uniaxial stress-strain diagram will indicate that the relationship between stress and strain is highly nonlinear.

Further testing will reveal history dependence, strong temperature dependence, and rate dependence in most structural materials. However, as shown in Fig. 8, there is a linear relation between the stress and the strain for a large portion of the load-carrying capability of the component. This range is called the linear elastic range, and the stress level at which this nonlinearity becomes significant is called the yield point. As a margin of safety against structural failure, it is customary to design most aerospace components such that at every point in the structure the yield point is not exceeded.

Most aerospace structural components are constructed using metals such as aluminum, stainless steel, titanium, and Inconel. For all practical purposes these materials are isotropic in their elastic range; that is, their properties at each material point are independent of the coordinate axis direction.

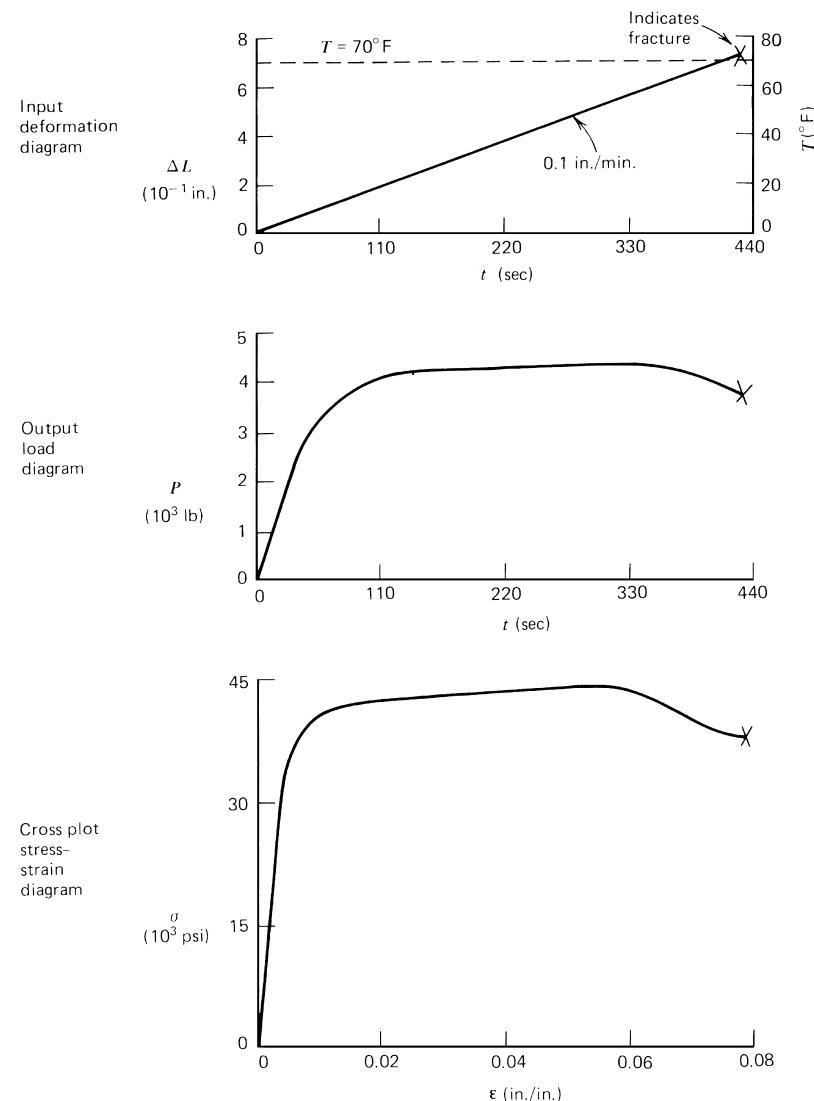


FIGURE 8 Uniaxial bar (Al 6061-T6) with cross-sectional area $A = 0.098 \text{ in.}^2$ subjected to monotonic deformation ΔL . (Results obtained on an Instron 1125 testing machine, Texas A&M University Mechanics and Materials Center.) [From Allen, D. H., and Haisler, W. E. (1985). "Introduction to Aerospace Structural Analysis," Wiley, New York.]

In the past two decades one of the most significant advances in aerospace structures has resulted from the development of composite materials. These materials are generally made of two or more materials, with one embedded in the other during some complex forming process. A wide variety of materials is currently under development. The embedding material may be continuous fibers, chopped fibers, or particulates. When fibers are used, they are most often made of graphite, boron, or fiberglass. Where particulates or chopped fibers are used, they may be boron, silicon carbide, or graphite. These materials are embedded in a matrix that is usually a polymer material such as epoxy. Metal matrix composites have been developed by utilizing materials such as aluminum for the matrix. Although

composite materials are very costly to fabricate compared to pure metals, their improved strength-to-weight ratios can make them cost effective when weight is a critical factor.

Materials of this type are being utilized in many critical areas on advanced tactical fighter aircraft, the shuttle and its companion launch vehicle, and even bicycle frames. Because of the orientation of the second-phase particles or fibers, these materials are inherently anisotropic.

1. Isotropic Elastic Materials

For isotropic elastic materials it can be shown by using symmetry constraints that

$$\begin{bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sigma_{yz} \\ \sigma_{xz} \\ \sigma_{xy} \end{bmatrix} = \frac{E}{(1+\nu)(1-2\nu)} \begin{bmatrix} 1-\nu & \nu & \nu & 0 & 0 & 0 \\ \nu & 1-\nu & \nu & 0 & 0 & 0 \\ \nu & \nu & 1-\nu & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1-2\nu}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1-2\nu}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1-2\nu}{2} \end{bmatrix} \begin{bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \\ \varepsilon_{yz} \\ \varepsilon_{xz} \\ \varepsilon_{xy} \end{bmatrix} - \frac{E\alpha(T-T_0)}{(1-2\nu)} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad (10)$$

where E , ν , and α are materials constants obtainable directly from two laboratory tests and T_0 is the reference temperature at which no stress is observed at zero strain.

2. Orthotropic Elastic Materials

Composite materials generally exhibit three mutually perpendicular planes of material symmetry. These materials are therefore called orthotropic. For orthotropic materials it can be shown by using symmetry constraints that

$$\begin{bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{23} \\ \sigma_{13} \\ \sigma_{12} \end{bmatrix} = \begin{bmatrix} D_{11} & D_{12} & D_{13} & 0 & 0 & 0 \\ D_{12} & D_{22} & D_{23} & 0 & 0 & 0 \\ D_{13} & D_{23} & D_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & D_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & D_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & D_{66} \end{bmatrix} \times \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \varepsilon_{33} \\ \varepsilon_{23} \\ \varepsilon_{13} \\ \varepsilon_{12} \end{bmatrix} + \begin{bmatrix} \beta_{11}(T-T_0) \\ \beta_{22}(T-T_0) \\ \beta_{33}(T-T_0) \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad (11)$$

where D_{ij} and β_{ij} are material constants that may be obtained from laboratory tests.

D. Field Problem Description

The field problem describing the analysis of aerospace structures can now be defined. There are 16 field variables to be determined at all points in the body of interest: σ_{xx} , σ_{yy} , σ_{zz} , σ_{yz} , σ_{xz} , σ_{xy} , ε_{xx} , ε_{yy} , ε_{zz} , ε_{yz} , ε_{xz} , ε_{xy} , u , v , w , and t . The governing field equations are six strain-displacement equations [Eqs. (5)], three conservation of momentum equations [Eqs. (6)], one conservation of energy equation [Eq. (9)], and six constitutive equations [Eqs. (10) for isotropic materials or Eqs. (11) for orthotropic materials]. Thus, there is a total of 16 equations in 16 unknowns. With the imposition of physically ad-

missable boundary conditions on the surface S , the problem is completely specified. The difficulty in obtaining solutions to problems of this type arises due to the fact that aerospace structures typically have extremely complex geometric shapes (boundaries).

The standard procedure for solving this problem is first to obtain the temperature field $T = T(x, y, z, t)$ by using the thermodynamic equation (9), since it is uncoupled from the remaining mechanical equations. The resulting temperature field may be used as input to solve the remaining 15 equations in 15 unknowns. For isothermal conditions the necessity to solve the heat transfer problem is obviated.

IV. ANALYSIS OF ADVANCED BEAMS

The geometric shape and/or loading conditions in the vast majority of structures are so complex that it is not possible to obtain exact solutions to the thermoelastic field problem. Therefore, a field of approximate techniques has been developed that is called strength of materials or structural mechanics. In these approximate methods it is generally assumed that the structural response is constrained by simplifying kinematic assumptions that are at least approximately correct. The resulting simplifications allow for the construction of approximate solutions for a wide variety of aerospace structures. One such simplified structure is the advanced beam. As shown in Fig. 9, the beam may be subjected to complex load conditions. In addition, the beam may be of heterogeneous material makeup, such as a composite or bonded component. Furthermore, the beam cross section normal to the longitudinal or x coordinate may be variable.

All of these complexities lead to very intricate beams that are nevertheless found in large space structures and in aircraft fuselages and empennages.

A. Equilibrium of Advanced Beams

Consider again Fig. 9. Suppose that a cutting plane is passed through the cross section, as shown, so that the

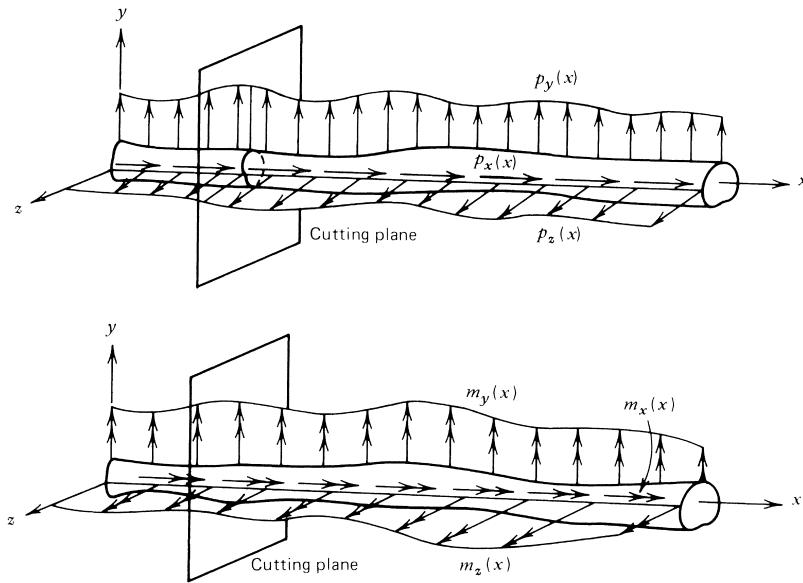


FIGURE 9 Advanced beam with externally applied forces and moments per unit length. [From Allen, D. H., and Haisler, W. E. (1985). "Introduction to Aerospace Structural Analysis," Wiley, New York.]

resulting free-body diagram is as shown in Fig. 10. We now define the following six resultants on any cross section:

$$P = P(x) \equiv \int_z \int_y \sigma_{xx} dy dz = \int_A \sigma_{xx} dA, \quad (12a)$$

$$V_y = V_y(x) \equiv \int_z \int_y \sigma_{xy} dy dz = \int_A \sigma_{xy} dA, \quad (12b)$$

$$V_z = V_z(x) \equiv \int_z \int_y \sigma_{xz} dy dz = \int_A \sigma_{xz} dA, \quad (12c)$$

$$M_x = M_x(x) \equiv \int_z \int_y \sigma_{xz}(y - \sigma_{xy}z) dy dz, \quad (12d)$$

$$M_y = M_y(x) \equiv \int_z \int_y \sigma_{xx}z dy dz = \int_A \sigma_{xx}z dA, \quad (12e)$$

$$M_z = M_z(x) \equiv -\int_z \int_y \sigma_{xx}y dy dz = -\int_A \sigma_{xx}y dA. \quad (12f)$$

It can be shown for quasi-static conditions ($\ddot{u} \approx 0$, $\ddot{v} \approx 0$, $\ddot{w} \approx 0$) by integrating the differential of equations of motion (6) over the cross section that

$$dP/dx = -p_x(x), \quad (13a)$$

$$dV_y/dx = -p_y(x), \quad (13b)$$

$$dV_z/dx = -p_z(x), \quad (13c)$$

$$dM_x/dx = -m_x(x), \quad (13d)$$

$$dM_y/dx = -m_y(x) + V_z(x), \quad (13e)$$

$$dM_z/dx = -m_z(x) - V_y(x), \quad (13f)$$

where the lowercase quantities on the right-hand side are input loads, as shown in Fig. 9. Thus, with appropriate boundary conditions, Eqs. (13) may be integrated to construct resultant load diagrams, which are statements of internal equilibrium at every point in the beam. By convention, $P = P(x)$ is called the axial load, $V_y = V_y(x)$ and $V_z = V_z(x)$ are called the shear loads, $M_x = M_x(x)$ is called the torque load, and $M_y = M_y(x)$ and $M_z = M_z(x)$ are called bending moments. The response of the beam to each of the resultants P , M_y , and M_z is detailed in the following sections. In many cases the response of advanced beams to the resultants V_y , V_z , and M_x is negligible. For that reason and in the interest of space, the reader is referred to the Bibliography for coverage of these resultants.

B. Bending and Extension of Advanced Beams

The subject of this section is the determination of the axial components of stress σ_{xx} and strain ϵ_{xx} due to the axial load P and bending moments M_y and M_z .

There are two important simplifying assumptions in the theory of advanced beams: (1) the transverse components of normal stress σ_{yy} and σ_{zz} are assumed to be negligible compared to the axial stress σ_{xx} , and (2) cross sections are assumed to remain planar and normal to the longitudinal axis of deformation, as shown in Fig. 11. The second assumption, called the Euler-Bernoulli assumption, may be represented mathematically by

$$u(x, y, z) = u_0 - \theta_z(x)y + \theta_y(x)z, \quad (14)$$

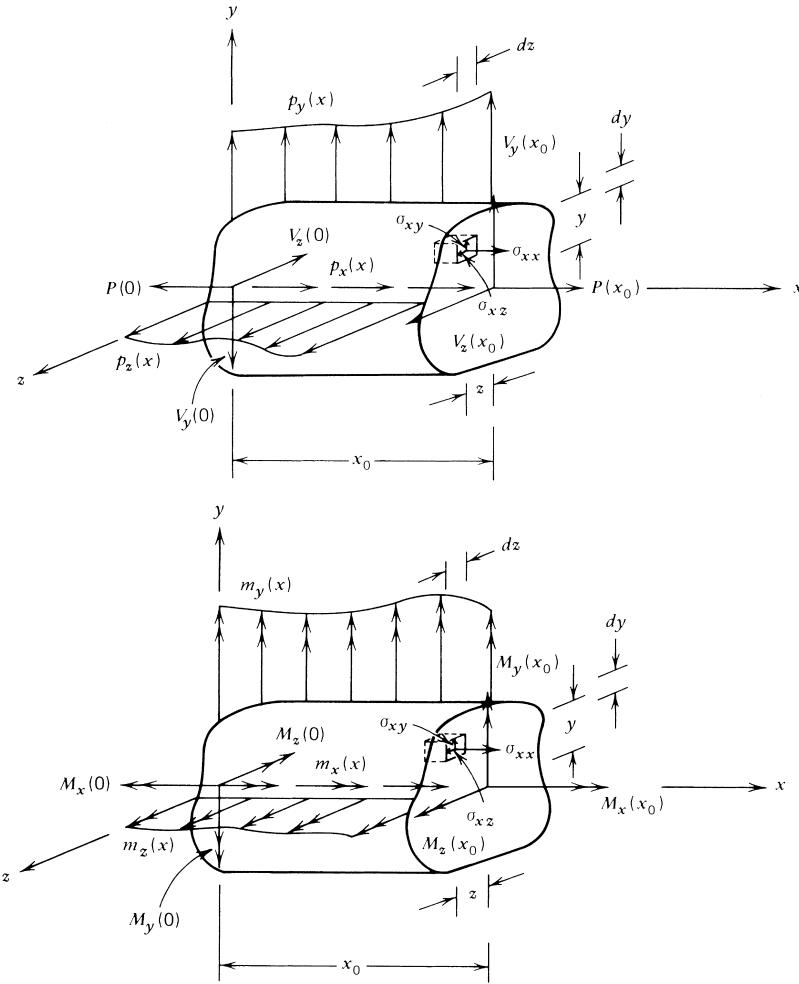


FIGURE 10 Free-body diagram of cut advanced beam. [From Allen, D. H., and Haisler, W. E. (1985). “Introduction to Aerospace Structural Analysis,” Wiley, New York.]

where u_0 is the midplane axial deformation and θ_z and θ_y are components of rotation of the cross section, as shown in Fig. 11. Substitution of Eq. (14) into the strain displacement relation (5), and this result into the stress-strain equation (11), and, finally, this result into the resultant load equations (12a), (12e), and (12f) will give, upon rearrangement,

$$\begin{aligned} \sigma_{xx} = & \frac{E(P + P^T)}{E_1 A^*} \\ & - \frac{E}{E_1} \left[\frac{(M_z - M_z^T) I_{yy}^* + (M_y + M_y^T) I_{yz}^*}{(I_{yy}^* I_{zz}^* - I_{yz}^{*2})} \right] y \\ & + \frac{1}{E_1} \left[\frac{(M_y + M_y^T) I_{zz}^* + (M_z - M_z^T) I_{yz}^*}{(I_{yy}^* I_{zz}^* - I_{yz}^{*2})} \right] z, \end{aligned} \quad (15)$$

where the following quantities are called modulus-weighted section properties:

$$A^* \equiv \int_A \frac{E}{E_1} dA, \quad (16a)$$

$$\bar{y}^* \equiv \frac{1}{A^*} \int_A \frac{E}{E_1} y dA, \quad (16b)$$

$$\bar{z}^* \equiv \frac{1}{A^*} \int_A \frac{E}{E_1} z dA, \quad (16c)$$

$$I_{yy}^* \equiv \int_A \frac{E}{E_1} z^2 dA, \quad (16d)$$

$$I_{yz}^* \equiv \int_A \frac{E}{E_1} yz dA, \quad (16e)$$

$$I_{zz}^* \equiv a | b \int_A \frac{E}{E_1} y^2 dA. \quad (16f)$$

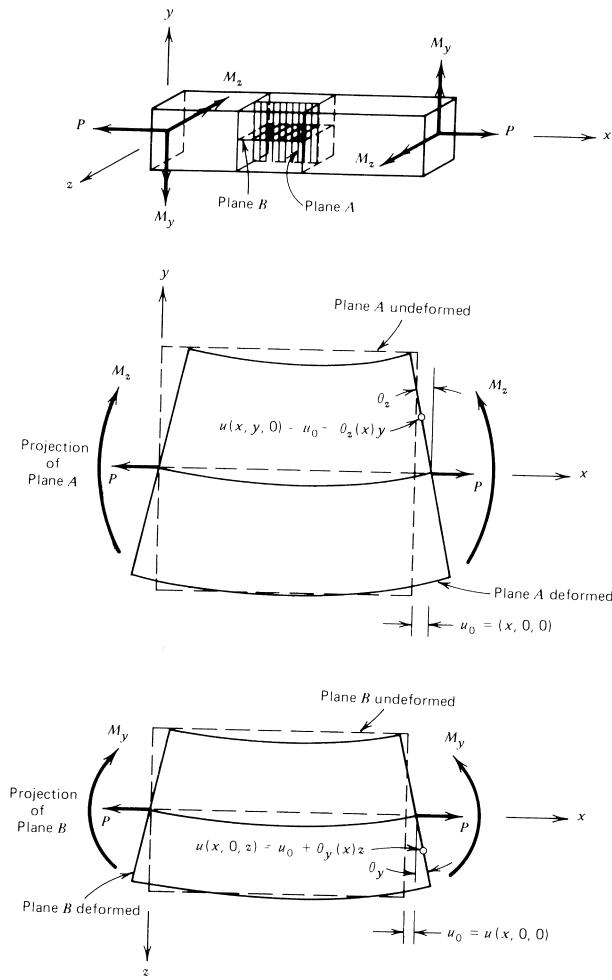


FIGURE 11 Kinematics of an advanced beam. [From Allen, D. H., and Haisler, W. E. (1985). “Introduction to Aerospace Structural Analysis,” Wiley, New York.]

Here E_1 is an arbitrary constant and A the area of the cross section normal to the x axis. Furthermore,

$$P^T \equiv \int_A E\alpha(T - T_0) dA, \quad (17a)$$

$$M_y^T \equiv \int E\alpha(T - T_0)z dA, \quad (17b)$$

$$M_z^T \equiv \int_A E\alpha(T - T_0)y dA. \quad (17c)$$

Equation (15) will give the axial stress component σ_{xx} at any point in the beam in terms of the resultants P , M_y , and M_z , which may be determined by using Eqs. (13).

It is possible to determine the three components of displacement at all points in an advanced beam. However, due to the Euler–Bernoulli assumption [Eq. (14)], it is of practical importance to determine only the three components of displacement of the centroidal axis; that is, $u_0(x) \equiv u(x, 0, 0)$, $v_0(x) \equiv v(x, 0, 0)$, and $w_0(x) \equiv w(x, 0, 0)$.

Using the Euler–Bernoulli assumption, it can be shown that for small deformations

$$\frac{d\theta_z}{dx} = \frac{d^2v_0}{dx^2}, \quad \frac{d\theta_y}{dx} = \frac{-d^2w_0}{dx^2}. \quad (18)$$

With the preceding equations it can be shown in the process of deriving Eq. (15) that

$$\frac{du_0}{dx} = \frac{P + P^T}{E_1 A^*}, \quad (19a)$$

$$\frac{d^2v_0}{dx^2} = \frac{(M_z - M_z^T)I_{yy}^* + (M_y + M_y^T)I_{yz}^*}{E_1(I_{yy}^* I_{zz}^* - I_{yz}^{*2})}, \quad (19b)$$

$$\frac{d^2w_0}{dx^2} = \frac{-(M_y + M_y^T)I_{zz}^* - (M_z - M_z^T)I_{yz}^*}{E_1(I_{yy}^* I_{zz}^* - I_{yz}^{*2})}. \quad (19c)$$

Since all of the quantities on the right-hand side of (19) can be determined from the external loads, geometry, material properties, and temperature field, they can be used as input to obtain the displacements u_0 , v_0 , and w_0 .

C. Example

Given: Suppose that the beam shown in Fig. 12 is a composite structure with properties as shown below.

Portion (i)	E_i (psi)	α_i (per °F)
1	30×10^6	5×10^{-6}
2	10×10^6	6.5×10^{-6}
3	10×10^6	6.5×10^{-6}
4	10×10^6	6.5×10^{-6}

Required:

- Determine modulus-weighted section properties about the modulus-weighted centroid.
- Transform the y' axis to the modulus-weighted centroid and construct resultant load diagrams.
- Determine the axial components of the stress and strain at the following points:
 - $(x, y, z) = (200, 2, 0)$,
 - $(x, y, z) = (200, -2, 0)$,
 - $(x, y, z) = (200, 0, 11.12)$, and
 - $(x, y, z) = (200, 0, -5.88)$,
 where the coordinate locations are measured with respect to the modulus-weighted axes.
- Rework part (c) assuming that, in addition to the mechanical loads, the beam is subjected to a thermal load $T - T_0 = 0.10x$ (°F), where x is in inches.
- Obtain the following displacements for the case where the blade is subjected to both thermal and mechanical loads:

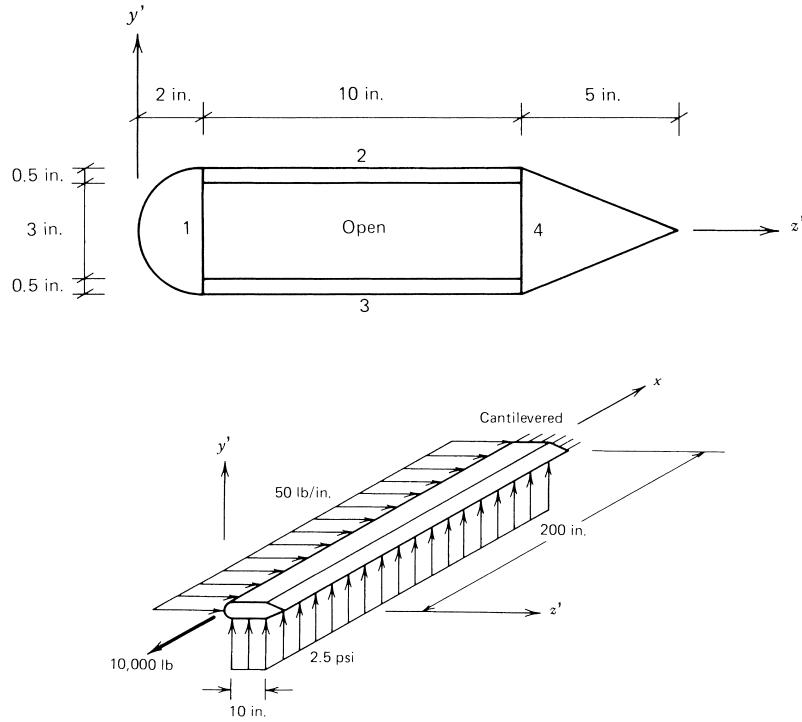


FIGURE 12 Composite helicopter rotor blade subjected to the loading shown. [From, Allen, D. H., and Haisler, W. E. (1985). "Introduction to Aerospace Structural Analysis," Wiley, New York.]

- (1) $u_0(x)$, $v_0(x)$, $w_0(x)$, and
- (2) $u_0(0)$, $v_0(0)$, $w_0(0)$.

Solution:

(a) By symmetry $\bar{y}' = 0$, $I_{yz}^* = 0$.

Now let $E_1 = 10 \times 10^6$ psi and tabulate Eq. (16c), as shown in [Table I](#). Now tabulate Eqs. (16e) and (16f) to find the modulus-weighted moments of inertia, as shown in [Table II](#).

(b) The applied loads must now be resolved to the modulus-weighted centroid. The statically equivalent

loading is shown in [Fig. 13](#). The internal resultants may be obtained from Eq. (13) as follows:

$$\begin{aligned} P(x) &= \cancel{P}(0) - \int_0^x \cancel{p}^0(x) dx \\ &\Rightarrow P(x) = 10,000 \text{ lb}, \\ V_y(x) &= \cancel{V}_y(0) - \int_0^x \cancel{p}_y^{25}(x) dx \\ &\Rightarrow V_y(x) = -25x \text{ lb}, \\ V_z(x) &= \cancel{V}_z(0) - \int_0^x \cancel{p}_z^{50}(x) dx \end{aligned}$$

TABLE I

Portion (<i>i</i>)	E_i (10^6 psi)	A_i (in. ²)	$\frac{E_i}{E_1} A_i$ (in. ²)	\bar{z}'_i (in.)	$\frac{E_i}{E_1} \bar{z}'_i A_i$ (in. ³)
1	30	6.28	18.84	1.15	21.67
2	10	5.00	5.00	7.00	35.00
3	10	5.00	5.00	7.00	35.00
4	10	10.00	10.00	13.67	136.70
$A^* = \sum \frac{E_i}{E_1} A_i = 38.84$			$\sum \frac{E_i}{E_1} \bar{z}'_i A_i = 228.37$		
$\bar{z}'^* = \frac{1}{A^*} \sum \frac{E_i}{E_1} \bar{z}'_i A_i = \frac{228.37}{38.84} \Rightarrow$			$\bar{z}'^* = 5.88 \text{ in.}$		

TABLE II

Portion (i)	A_i (in. ²)	E_i (10 ⁶ psi)	\bar{y}'_i (in.)	\bar{z}'_i (in.)	$I_{y_0 y_{0i}}$ (in. ⁴)	$I_{z_0 z_{0i}}$ (in. ⁴)	$\frac{E_i}{E_1} (I_{y_0 y_{0i}} + \bar{z}'_i^2 A_i)$ (in. ⁴)	$\frac{E_i}{E_1} (I_{z_0 z_{0i}} + \bar{y}'_i^2 A_i)$ (in. ⁴)
1	6.28	30	0.00	1.15	1.76	6.28	30.21	18.25
2	5.00	10	1.75	7.00	41.67	0.10	286.67	15.42
3	5.00	10	-1.75	7.00	41.67	0.10	286.67	15.42
4	10.00	10	0.00	13.67	13.89	6.67	<u>1882.58</u>	<u>6.67</u>
						$I_{y'y'}^* = 2486.13$	$I_{z'z'}^* = 56.36$	
						$I_{yy}^* = I_{y'y'}^* - (\bar{z}'^*)^2 A^* = 2486.13 - 5.88^2 \cdot 38.84 \Rightarrow$		
						$I_{yy}^* = 1143.3 \text{ in.}^4$		
						$I_{zz}^* = I_{z'z'}^* - (\bar{y}'^*)^2 A^* = 56.36 - 0^2 \cdot 38.84 \Rightarrow$		
						$I_{zz}^* = 56.36 \text{ in.}^4$		

$$\Rightarrow V_z(x) = -50x \text{ lb},$$

$$M_x(x) = M_x^0(0) - \int_0^x \bar{m}_x^{28}(x) dx$$

$$\Rightarrow M_x(x) = 28x \text{ in. lb},$$

$$M_y(x) = M_y^{58,800}(0) - \int_0^x \left[\bar{m}_y^0(0) - \bar{M}_z^{50\times}(x) \right] dx$$

$$\Rightarrow M_y(x) = -58,800 - 25x^2 \text{ in. lb},$$

$$M_z(x) = M_z^0(0) - \int_0^x \left[\bar{m}_z^0(x) + \bar{M}_y^{25\times}(x) \right] dx$$

$$\Rightarrow M_z(x) = 12.5x^2 \text{ in. lb},$$

where x is measured in inches. This is shown graphically in Fig. 14.

(c) Utilizing Eqs. (10) and (15) gives

$$\varepsilon_{xx}(x, y, z) = \frac{P}{E_1 A^*} - \frac{M_z y}{E_1 I_{zz}^*} + \frac{M_y z}{E_1 I_{yy}^*} \Rightarrow,$$

$$\begin{aligned} \varepsilon_{xx}(200, 2, 0) &= \frac{10,000}{10^7 \cdot 38.84} - \frac{500,000 \cdot 2}{10^7 \cdot 56.36} \\ &\quad + \frac{(-1,058,800) \cdot 0}{10^7 \cdot 1143.3} \Rightarrow, \end{aligned}$$

$$\varepsilon_{xx}(200, 2, 0) = -0.00175 \text{ in./in.},$$

$$\begin{aligned} \sigma_{xx} &= E \varepsilon_{xx} \Rightarrow \sigma_{xx}(200, 2, 0) \\ &= 10^7 \cdot (-0.00175) \Rightarrow, \end{aligned}$$

$$\sigma_{xx}(200, 2, 0) = -17,500 \text{ psi.}$$

Similarly,

$$\varepsilon_{xx}(200, -2, 0) = 0.00180 \text{ in./in.},$$

$$\sigma_{xx}(200, -2, 0) = 18,000 \text{ psi},$$

$$\varepsilon_{xx}(200, 0, 11.12) = -0.00100 \text{ in./in.},$$

$$\sigma_{xx}(200, 0, 11.12) = -10,000 \text{ psi},$$

$$\varepsilon_{xx}(200, 0, -5.88) = 0.000570 \text{ in./in.},$$

$$\sigma_{xx}(200, 0, -5.88) = 17,100 \text{ psi.}$$

(d) Next,

$$P^T = \int_A E \alpha \Delta T dA = \sum_{i=1}^4 E_i \alpha_i A_i \Delta T,$$

$$M_y^T = \int_A E \alpha \Delta T z dA = \sum_{i=1}^4 E_i \alpha_i \bar{z} A_i \Delta T,$$

$$M_z^T = \int_A E \alpha \Delta T y dA = \sum_{i=1}^4 E_i \alpha_i \bar{y}_i A_i \Delta T.$$

Table III presents this in tabular form.

Utilizing Eqs. (10) and (15) gives

$$\begin{aligned} \varepsilon_{xx}(x, y, z) &= \frac{(P + P^T)}{E_1 A^*} - \frac{1}{E_1} \left(\frac{M_z - M_z^T}{I_{zz}^*} \right) y \\ &\quad + \frac{1}{E_1} \left(\frac{M_y + M_y^T}{I_{yy}^*} \right) z \Rightarrow, \end{aligned}$$

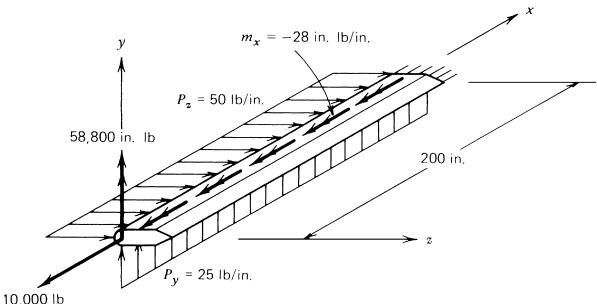


FIGURE 13 External loading configuration for composite beam. [From Allen, D. H., and Haisler, W. E. (1985). "Introduction to Aerospace Structural Analysis," Wiley, New York.]

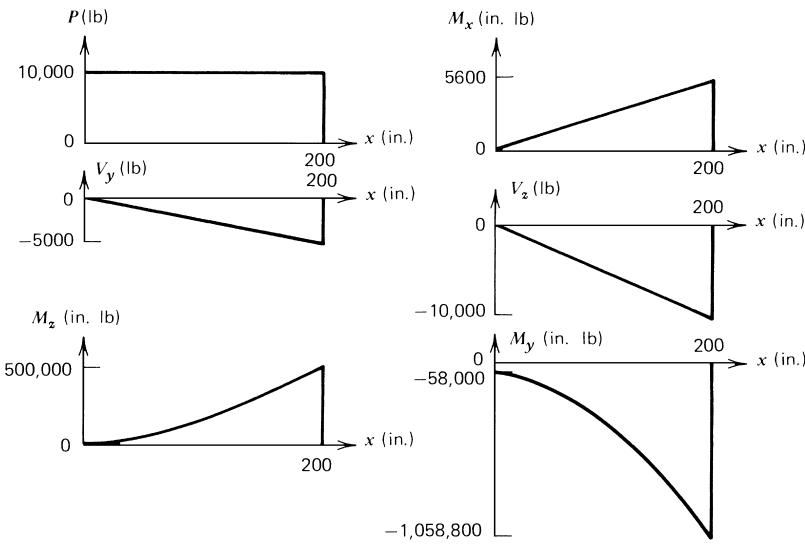


FIGURE 14 Resultant load diagrams for composite beam. [From Allen, D. H., and Haisler, W. E. (1985). "Introduction to Aerospace Structural Analysis," Wiley, New York.]

$$\begin{aligned}\varepsilon_{xx}(200, 2, 0) &= \frac{(10,000 + 224.2 \cdot 200)}{10^7 \cdot 38.84} \\ &\quad - \frac{(500,000 - 0) \cdot 2}{10^7 \cdot 56.36} \\ &\quad + \frac{(-1,058,800 + 133.6 \cdot 200) \cdot 0}{10^7 \cdot 1143.3} \Rightarrow,\end{aligned}$$

$$\varepsilon_{xx}(200, 2, 0) = -0.00163 \text{ in./in.},$$

$$\sigma_{xx} = E(\varepsilon_{xx} - \alpha \Delta T) \Rightarrow,$$

$$\begin{aligned}\sigma_{xx}(200, 2, 0) &= 10^7 \cdot (-0.00163) \\ &\quad - 6.5 \cdot 10^{-6} \cdot 0.10 \cdot 200 \Rightarrow,\end{aligned}$$

$$\sigma_{xx}(200, 2, 0) = -17,600 \text{ psi}.$$

Similarly,

$$\varepsilon_{xx}(200, -2, 0) = 0.00192 \text{ in./in.},$$

$$\sigma_{xx}(200, -2, 0) = 17,900 \text{ psi},$$

$$\varepsilon_{xx}(200, 0, 11.12) = -0.000863 \text{ in./in.},$$

$$\sigma_{xx}(200, 0, 11.12) = -9930 \text{ psi},$$

$$\varepsilon_{xx}(200, 0, -5.88) = 0.000672 \text{ in./in.},$$

$$\sigma_{xx}(200, 0, -5.88) = 5420 \text{ psi}.$$

(e) For this case Eq. (19) reduces to

$$\begin{aligned}\frac{du_0}{dx} &= \frac{P + P^T}{E_1 A^*} = \frac{10,000 + 224.2x}{10^7 \cdot 38.84} \\ &= 2.575 \cdot 10^{-5} + 5.772 \cdot 10^{-7}x,\end{aligned}$$

TABLE III

Portion (<i>i</i>)	<i>A_i</i> (in. ²)	<i>E_i</i> (10 ⁶ psi)	<i>α_i</i> (10 ⁻⁶ /°F)	Δ <i>T</i> (°F)	<i>E_iα_iA_iΔT</i> (lb)
1	6.28	30	5.0	0.10 <i>x</i>	94.2 <i>x</i>
2	5.00	10	6.5	0.10 <i>x</i>	32.5 <i>x</i>
3	5.00	10	6.5	0.10 <i>x</i>	32.5 <i>x</i>
4	10.00	10	6.5	0.10 <i>x</i>	<u>65.0<i>x</i></u>
					<u>P^T = 224.2<i>x</i></u>
<i>z̄_i</i> (in.)	<i>E_iα_iz̄_iA_iΔT</i> (in. lb)	<i>ȳ_i</i> (in.)	<i>E_iα_iȳ_iA_iΔT</i> (in. lb)		
-4.73	-445.6 <i>x</i>	0.00	0.0		
1.12	36.4 <i>x</i>	1.75	56.9 <i>x</i>		
1.12	36.4 <i>x</i>	-1.75	-56.9 <i>x</i>		
7.79	<u>506.4<i>x</i></u>	0.00	<u>0.0</u>		
	<u>M_y^T = 133.6<i>x</i></u>		<u>M_z^T = 0.0</u>		

$$\begin{aligned}\frac{d^2v_0}{dx^2} &= \frac{(M_z - M_z^T)}{EI_{zz}^*} = \frac{(12.5x^2 - 0)}{10^7 \cdot 56.36} \\ &= 2.218 \cdot 10^{-8}x^2, \\ \frac{d^2w_0}{dx^2} &= \frac{-(M_y + M_y^T)}{EI_{yy}^*} \\ &= -\frac{(-81400 - 25x^2 + 133.6x)}{10^7 \cdot 1143.3} \\ &= 7.120 \cdot 10^{-6} - 1.169 \cdot 10^{-8}x + 2.187 \cdot 10^{-9}x^2.\end{aligned}$$

Integrating the preceding equations and utilizing the boundary conditions described in part (a) will result in

$$\begin{aligned}u_0(x) &= 2.886 \cdot 10^{-7}x^2 + 2.575 \cdot 10^{-5}x - 1.669 \cdot 10^{-2}, \\ v_0(x) &= 1.848 \cdot 10^{-9}x^4 - 5.915 \cdot 10^{-2}x + 8.873, \\ w_0(x) &= 1.822 \cdot 10^{-10}x^4 - 1.948 \cdot 10^{-9}x^3 \\ &\quad + 3.560 \cdot 10^{-6}x^2 - 7.022 \cdot 10^{-3}x + 0.986.\end{aligned}$$

Therefore,

$$\begin{aligned}u_0(0) &= -1.669 \cdot 10^{-2} \text{ in.,} \\ v_0(0) &= 8.873 \text{ in.,} \\ w_0(0) &= 0.986 \text{ in.}\end{aligned}$$

The preceding stresses and displacements may be checked against design constraints in order to determine whether the structure design is acceptable.

V. ADVANCED ANALYSIS TECHNIQUES

The analysis techniques described in the previous section can be applied to cantilever beams such as wings and fuselages because they are statically determinate structures. However, for more detailed analysis of aerospace structures, the input resultant loads described in Eq. (10) cannot be determined a priori because the structures are statically indeterminate. For these structures more complex analysis techniques have been developed. The most powerful and commonly used technique for statically indeterminate structures involves the use of variational or energy methods.

Variational methods typically convert all of the pointwise governing differential equations for the problem at hand into a single global equation for the entire body that is amenable to approximation and has the capability to handle complex boundary conditions.

The advent of the high-speed digital computer has brought the capability to perform structural analysis on extremely complicated vehicles, such as that shown in Fig. 1. In almost all cases the variational equations governing structural response have been discretized using the

finite element method. This method utilizes assumed solution forms over small subsets of the structure of interest to obtain an accurate approximation of the stresses, strains, and displacements throughout the structure. The choice of the assumed solution form results in a set of coupled algebraic equations that are particularly amenable to solution on a large digital computer.

Because of the power of variational methods and the finite element method, they are reviewed in some detail in the following sections.

A. Variational Methods

Although variational methods may be utilized to solve extremely complex problems, for the purpose of demonstration, a simple example is chosen. Consider a homogeneous beam at a constant temperature and subjected to bending in the $x-z$ plane (see Fig. 10). For this special case M_z is the only nonzero resultant load. Therefore, the displacement equation (19c) simplifies to

$$\frac{d^2w_0}{dx^2} = -M_y/EI_{yy}. \quad (20)$$

Substitution of the preceding into differential equations of equilibrium [(13b) and (13f)] therefore results in (assuming $m_z = 0$)

$$\frac{d^2}{dx^2} \left(EI_{yy} \frac{d^2w_0}{dx^2} \right) - p_z(x) = 0. \quad (21)$$

The preceding equation can be utilized to obtain the beam transverse displacement field $w_0 = w_0(x)$ in terms of the applied loading $p_z(x)$.

In many cases it is possible to integrate the preceding differential equation directly to obtain an exact solution. However, suppose that this cannot be accomplished. One alternative is to construct a variational principle. To do this, first multiply Eq. (21) against a small but arbitrary variation in the displacement field $\delta w_0 = \delta w_0(x)$ and integrate over the length L of the beam to obtain

$$\int_0^L \left[\frac{d^2}{dx^2} \left(EI_{yy} \frac{d^2w_0}{dx^2} \right) - p_z(x) \right] \delta w_0(x) dx = 0. \quad (22)$$

The preceding is called a variational principle. It has the property that any solution to the pointwise differential equation (21) is also a solution to the global equation (22). In most structural analyses it is customary to obtain a weakened form of (3) by integrating by parts twice to obtain

$$\begin{aligned}& \int_0^L EI_{yy} \frac{d^2w_0}{dx^2} \delta \frac{d^2w_0}{dx^2} dx \\ &= \int_0^L p_z \delta w_0 dx - V_z(0) \delta w_0(0) - M_y(0) \delta \frac{dw_0}{dx}(0) \\ &\quad + V_z(L) \delta w_0(L) + M_y(L) \delta \frac{dw_0}{dx}(L),\end{aligned} \quad (23)$$

where the equilibrium equations (12) have been used to obtain the boundary conditions. [The resulting equation is called weakened because less smoothness is required of the dependent variable $w_0(x)$.] It can be shown that the left side of Eq. (23) represents the internal work (δW_i) done on the structure, and the right side represents the external work (δW_e) due to the virtual displacement field $\delta w_0(x)$. Thus, Eq. (23) is also an energy equation, often called the principle of virtual work, and often written

$$\delta W = \delta W_i + \delta W_e = 0. \quad (24)$$

Because Eq. (23) is a global statement that explicitly contains the boundary conditions, it is particularly amenable to approximate analysis techniques. One technique, due to Lord Rayleigh (1842–1919), is called the Rayleigh–Ritz method. To see how this method may be used, consider the cantilever beam subjected to an evenly distributed load p_0 shown in Fig. 15.

We take as an initial approximation to $w_0(x)$ the quadratic polynomial

$$w_0(x) = \bar{a}_1 + \bar{a}_2x + \bar{a}_3x^2.$$

Because the beam is cantilevered at $x = 0$, the geometric boundary conditions are

$$w_0(0) = dw_0(0)/dx = 0.$$

Satisfaction of these geometric boundary conditions requires that

$$\bar{a}_1 = \bar{a}_2 = 0.$$

Hence, $w_0(x)$ becomes

$$w_0(x) = a_1x^2,$$

where, for convenience, \bar{a}_3 is redefined to be a_1 . We must choose a virtual displacement field $\delta w_0(x)$. It has been noted previously that the virtual displacement is actually a variation or differential of the actual displacement. Hence, we write

$$\delta w_0(x) = (\delta a_1)x^2,$$

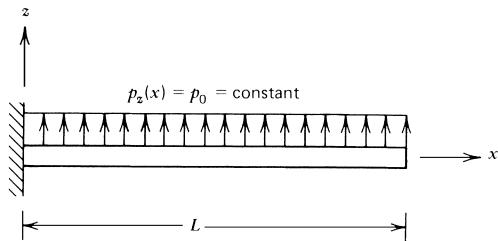


FIGURE 15 Cantilever beam subjected to an evenly distributed load p_0 . [From Allen, D. H., and Haisler, W. E. (1985). “Introduction to Aerospace Structural Analysis,” Wiley, New York.]

which is consistent with the geometry constraints; that is, $\delta w_0(x)$ satisfies the boundary conditions. Since by inspection $V_z(L) = M_y(L) = 0$, the total virtual work becomes

$$\begin{aligned} \delta W &= 0 \\ &= - \int_0^L EI_{yy} \frac{d^2 w_0}{dx^2} \frac{d^2(\delta w_0)}{dx^2} dx + \int_0^L p_z \delta w_0 dx \\ &= - \int_0^L EI_{yy} (2a_1)(2\delta a_1) dx + \int_0^L p_0(\delta a_1)x^2 dx \\ &= -4EI_{yy}La_1 \delta a_1 + (p_0L^3/3) \delta a_1 \\ &= (-4EI_{yy}La_1 + p_0L^3/3) \delta a_1. \end{aligned}$$

For the virtual displacement to be arbitrary, $\delta a_1 \neq 0$, and hence

$$a_1 = p_0L^2/12EI_{yy}.$$

Therefore,

$$w_0(x) = p_0L^2/12EI_{yy}.$$

Evaluating $w_0(x)$ at $x = 0.75L$, we find

$$w_0(0.75L) = 0.0469(p_0L^4/EI_{yy}).$$

This can be compared to the exact displacement, which can be obtained by double integrating the governing differential equation,

$$w_0(0.75L)_{\text{exact}} = 0.0835(p_0L^4/EI_{yy}),$$

which indicates an error of 44% at $x = 0.75L$.

Suppose we take the two-term approximation

$$w_0(x) = a_1x^2 + a_2x^3$$

in order to improve our approximate results. The discretized virtual work expression becomes

$$\begin{aligned} \delta W &= 0 \\ &= - \int_0^L EI_{yy} (2a_1 + 6a_2x)(2\delta a_1 + 6\delta a_2x) dx \\ &\quad + \int_0^L p_0(\delta a_1x^2 + \delta a_2x^3) dx \\ &= \left[-EI_{yy}(4La_1 + 6L^2a_2) + \frac{p_0L^3}{3} \right] \delta a_1 \\ &\quad + \left[-EI_{yy}(6L^2a_1 + 12L^3a_2) + \frac{p_0L^4}{4} \right] \delta a_2. \end{aligned}$$

Requiring δa_1 and δa_2 to be arbitrary gives the two equations

$$\begin{bmatrix} 4EI_{yy}L & 6EI_{yy}L^2 \\ 6EI_{yy}L^2 & 12EI_{yy}L^3 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} p_0L^3/3 \\ p_0L^4/4 \end{bmatrix}$$

with the solution

$$\begin{aligned} a_1 &= \frac{5}{24}(p_0 L^2 / EI_{yy}), \\ a_2 &= -\frac{1}{12}(p_0 L / EI_{yy}). \end{aligned}$$

Thus, $w_0(x)$ is given by

$$w_0(x) = (p_0/24EI_{yy})(5L^2x^2 - 2Lx^3).$$

At $x = 0.75L$, the preceding equation gives

$$w_0(0.75L) = 0.08203(p_0 L^4 / EI_{yy}),$$

which is in error by approximately 2%.

For the three-term approximation

$$w_0(x) = a_1 x^2 + a_1 x^3 + a_4 x^4,$$

we obtain

$$w_0(x) = (p_0/24EI)(6L^2x^2 - 4Lx^3 + x^4),$$

which is the exact solution. If additional terms are added to $w_0(x)$, the solution remains the same.

It is instructive to compare displacements and bending moments given by the three approximations with the exact solution. We know that $M_y = -EI_{yy}(d^2 w_0 / dx^2)$ from strength of materials. **Table IV** shows the results.

Because an exact solution exists for the preceding problem, it is not necessary to use approximate methods such as the Rayleigh–Ritz. However, the power of this technique is demonstrated by the example. Furthermore, in real aerospace structures, $p_z(x)$ and $I_{yy}(x)$ may be complicated functions, and the structure may be statically indeterminate. For these applications variational methods may be the most powerful tool to obtain approximate results.

In the case of a beam subjected to planar bending, it has been shown that the governing differential equations resulted in the variational principle given by Eq. (23). For more complex problems similar variational principles may be constructed. For example, for an isotropic elastic three-dimensional body, Eqs. (5)–(8) may be combined to give

$$\delta W = \delta W_i + \delta W_e = 0, \quad (25)$$

where

$$\begin{aligned} \delta W_i &= - \int_V (\sigma_{xx} \delta \varepsilon_{xx} + \sigma_{yy} \delta \varepsilon_{yy} + \sigma_{zz} \delta \varepsilon_{zz} + 2\sigma_{yz} \delta \varepsilon_{yz} \\ &\quad + 2\sigma_{xz} \delta \varepsilon_{xz} + 2\sigma_{xy} \delta \varepsilon_{xy}) dV \end{aligned} \quad (26)$$

and

$$\begin{aligned} \delta W_e &= \int_S (T_x \delta u + T_y \delta v + T_z \delta w) dS \\ &\quad + \int_V (X \delta u + Y \delta v + Z \delta w) dV \end{aligned} \quad (27)$$

and the body forces X , Y , and Z may contain inertial effects for dynamic applications.

It can be seen that the general form of Eqs. (25)–(27) is similar to Eqs. (23) and (24). Furthermore, the Rayleigh–Ritz method may be used in conjunction with Eqs. (25)–(27) to obtain approximate solutions for three-dimensional aerospace structures.

B. The Finite Element Method

For complex structures the Rayleigh–Ritz procedure may break down owing to the fact that sufficient accuracy can be obtained only with an insoluble assumed solution form. An alternative method has therefore evolved in the twentieth century. Although initially called the stiffness method, this method was dubbed the finite element method (FEM) in 1960.

In the FEM, as in the Rayleigh–Ritz procedure, an assumed solution form is used with the variational principle (or some other similar global equation). However, in contrast to the Rayleigh–Ritz method, the FEM utilizes the assumed solution over a subdomain of the structure called a finite element. The approximate equations for this one element are then assembled with the governing equations for all of the other elements to obtain an approximate solution for the entire structure. Because of the way in which the element properties are assumed, the resulting equations are particularly amenable to solution on a large-core mainframe digital computer. Furthermore, the observed

TABLE IV

$w_0(x)$	$w_0(0.75L) \left(\frac{EI_{yy}}{p_0 L^4} \right)$	$M_y(x)$	$M_y(0)$	$M_y(L)$
One term	0.0469	$\frac{1}{6} p_0 L^2$	$\frac{1}{6} p_0 L^2$	$\frac{1}{6} p_0 L^2$
Two terms	0.0820	$\frac{p_0 L}{12}(5L - 6x)$	$\frac{5}{12} p_0 L^2$	$\frac{-1}{12} p_0 L^2$
Three terms	0.0835	$\frac{p_0}{2}(L - x)^2$	$\frac{1}{2} p_0 L^2$	0
Exact	0.0835	$\frac{p_0}{2}(L - x)^2$	$\frac{1}{2} p_0 L^2$	0

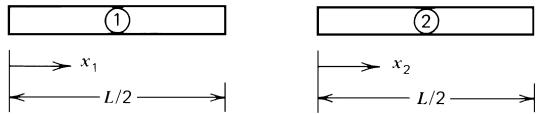


FIGURE 16 Two finite elements used to approximate the beam shown in Fig. 12. [From Allen, D. H., and Haisler, W. E. (1985). “Introduction to Aerospace Structural Analysis,” Wiley, New York.]

kinematics of various components can be used to construct highly accurate elements for the various subcomponents of the structure. For example, the highly complex large space structure shown in Fig. 2 might be modeled as an assemblage of beam elements for the trusslike appendages, connected to plate elements representing the solar collectors, and rigid bodies approximating the crew compartments.

To see how the FEM may be employed, consider again the beam example problem shown in Fig. 15. Suppose, however, that in the current section the variational principle is applied twice, each time over one-half the length of the beam ($0 \leq x_1 \leq L/2$ and $L/2 \leq x_2 \leq L$), as shown in Fig. 16.

If a quadratic displacement field is assumed over each subdomain, the result of applying the variational principle [Eq. (23)] over each range and solving for the resulting unknowns will be

$$w_0(x_1) = \frac{7}{48} \frac{p_0 L^2}{EI_{yy}} x_1^2,$$

$$w_0(x_2) = \frac{7}{192} \frac{p_0 L^4}{EI_{yy}} + \frac{7}{48} \frac{p_0 L^3}{EI_{yy}} x_2 + \frac{1}{48} \frac{p_0 L^2}{EI_{yy}} x_2^2.$$

It is instructive to compare the accuracy of the one- and two-element approximations. For example, evaluating the deflection at $x = 0.75L$ (three-quarter span point) yields the following.

$w_0(x)$ approximation	$\frac{EI_{yy}}{p_0 L^4}$	$M_y(0)$
One-element, quadratic	0.0469	$0.167 p_0 L^2$
One-element, cubic	0.0820	$0.417 p_0 L^2$
One-element, quartic	0.0835	$0.5 p_0 L^2$
Two-element, quadratic	0.0742	$0.292 p_0 L^2$
Exact	0.0835	$0.5 p_0 L^2$

The preceding results indicate that when the simple quadratic approximation is used, the solution improves significantly when two elements are used instead of one. One would expect that an idealization containing additional (shorter) elements would improve the results and that, in the limit as the number of elements is increased, the solution approaches the exact one. Herein lies the inherent strength of the FEM. Through so-called mesh refinement, either increasing the number of elements or redistribut-

ing them, the FEM can be utilized to obtain virtually any desired degree of accuracy (subject to computational cost constraints) even though an exact solution cannot be obtained. For example, analysis can be performed using brick elements and variational principle [(25)] on the complex fitting shown in Fig. 17.

The FEM is now an indispensable analysis tool in every structural design group in the aerospace industry and the National Aeronautics and Space Administration. Numerous extremely large and flexible computer codes such as NAS-TRAN and ANSYS now are available for analyzing such complex structures as the F-14 shown in Fig. 1. These computer codes are endowed with many user-friendly capabilities such as internal mesh optimization, banded equation solvers, and extremely complex boundary conditions. In addition, they contain large libraries of element types, including plate, shell, and higher-order continuum elements. Often they are capable of solving coupled thermomechanical problems, as well as handling material and geometric nonlinearities.

It must be remembered that the analytic capability of these computer codes is only as useful as the input

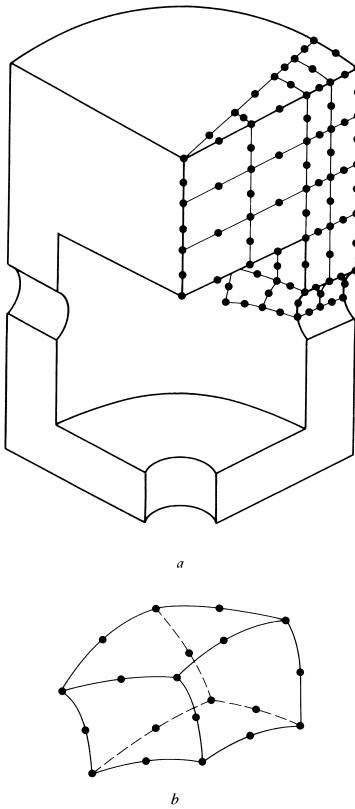


FIGURE 17 Three-dimensional structure modeled with solid brick elements. [From Allen, D. H., and Haisler, W. E. (1985). “Introduction to Aerospace Structural Analysis,” Wiley, New York.]

assumptions. Since all analytical tools in general entail a certain number of simplifying assumptions, the results of these analyses must always be compared against independently obtained experimental results. However, with accurate analysis tools it is possible to reduce the margins of safety and thus achieve weight savings, the critical constraint in aerospace structures.

C. Life Prediction Methodologies

Until recently it was believed that the primary structure for flight vehicles could be designed to produce essentially infinite life by limiting the design stresses so that some allowable stress is never exceeded in the structure. However, several commercial airline crashes in the last half of the 1980s have forced the structural analysis community to revise its approach to stress analysis. It was found that these aircraft underwent long-term fatigue-induced damage, which eventually lead to major structural damage. This damage can be exacerbated by adverse environmental conditions such as saltwater surroundings. This is due primarily to the fact that the joining of dissimilar structural parts such as thin skins and stringers causes stress concentrations that can induce microcracks that can grow as a result of long-term fatigue and eventually cause rupture of the part and failure of the entire air- or spacecraft.

As shown in Fig. 18, engineers are now developing a methodology that implements a fatigue crack growth prediction model to aerospace structure finite element computer codes. Although this procedure is still undergoing development at this time, it has reached a fairly mature state for metals. Considerably more research will be nec-

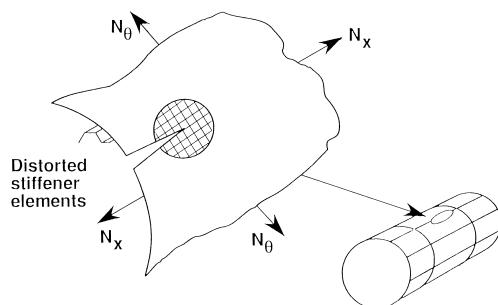


FIGURE 18 Global/local analysis scheme for a fuselage structure with an imbedded crack.

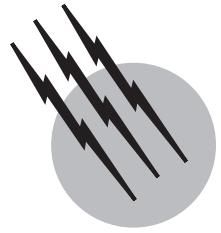
essary before accurate life predictions can be made for composite media.

SEE ALSO THE FOLLOWING ARTICLES

AIRCRAFT PERFORMANCE AND DESIGN • COMPOSITE MATERIALS • COMPUTATIONAL AERODYNAMICS • FLIGHT (AERODYNAMICS) • FRACTURE AND FATIGUE • MECHANICS OF STRUCTURES • RELIABILITY THEORY

BIBLIOGRAPHY

- Allen, D. H., and Haisler, W. E. (1985). "Introduction to Aerospace Structural Analysis," Wiley, New York.
- Bathe, K. J. (1982). "Finite Element Procedures in Engineering Analysis," Prentice-Hall, Englewood Cliffs, NJ.
- Beer, F. P., and Johnston, E. R., Jr. (1977). "Vector Mechanics for Engineers Statics and Dynamics," 3rd ed. McGraw-Hill, New York.
- Bruhn, E. F. (1965). "Analysis and Design of Flight Vehicle Structures," Tri-State Offset, Cincinnati, OH.
- Harris, C. E. (1990). "NASA Airframe Structural Integrity Program," NASA TM 102637.
- Jones, R. M. (1975). "Mechanics of Composite Materials," McGraw-Hill, New York.
- Lo, D. C., Allen, D. H., and Harris, C. E. (1990). In "Inelastic Deformation of Composite Materials" (G. J. Dvorak, ed.). Springer-Verlag, New York.
- Malvern, L. E. (1967). "Introduction to the Mechanics of a Continuous Medium," Prentice-Hall, Englewood Cliffs, NJ.
- Meriam, J. L. (1978). "Engineering Mechanics Statics and Dynamics," Wiley, New York.
- "Metallic Materials and Elements for Flight Vehicle Structures" (1976). Military Standardization Handbook, MIL-HDBK-5C, Sept.
- Newman, J. C., Jr. (1981). In "Methods and Models for Predicting Fatigue Crack Growth Under Random Loadings," ASTM STP 748, pp. 53–84, American Society for Testing and Materials.
- Oden, J. T., and Ripperger, A. E. (1981). "Mechanics of Elastic Structures," 2nd ed., Hemisphere, Washington, DC.
- Peery, D. J. (1950). "Aircraft Structures," McGraw-Hill, New York.
- Peery, D. J., and Azar, J. J. (1982). "Aircraft Structures," 2nd ed., McGraw-Hill, New York.
- Reddy, J. N. (1984). "An Introduction to the Finite Element Method," McGraw-Hill, New York.
- Reismann, H., and Pawlik, P. S. (1980). "Elasticity Theory and Applications," Wiley, New York.
- Rivello, R. M. (1969). "Theory and Analysis of Flight Structures," McGraw-Hill, New York.
- Timoshenko, S. P., and Goodier, J. N. (1970). "Theory of Elasticity," 3rd ed., McGraw-Hill, New York.
- Zienkiewicz, O. C. (1977). "The Finite Element Method," McGraw-Hill, New York.



Ramjets and Scramjets

Bernard Petit

ONERA (*France*)

- I. History and Possible Applications
- II. Ramjet Operation
- III. Integrated Booster
- IV. Liquid Fuelled Ramjet
- V. Solid Fuelled Ramjet
- VI. Scramjet
- VII. Development Facilities

GLOSSARY

A Local cross section.

Atmosphere See the variation of the static temperature (T_0) and the static pressure (p_0) in Fig. 1 as a function of the altitude and climate. At sea level, in standard atmosphere, the temperature is 288.15 K and the pressure is 101,325 Pa.

Fuel controlled ducted rocket Ramrocket whose gas generator is separated from the combustor by a variable sonic section.

γ Specific heats ratio ($\gamma = C_p/C_v$).

m Mass.

\dot{m} Mass flow rate.

Mach number (M) Ratio of the speed (V) to the local speed of sound (a) which depends on the fluid (by γ) and the local static temperature (T) ($M = V/a = V/\sqrt{\gamma \cdot R \cdot T}$). At sea level, in standard atmosphere, the speed of sound is 340.29 m/s.

Ramrocket Solid fuelled ramjet.

“Rustique” ducted rocket Ramrocket whose gas generator is separated from the combustor by a fixed non-

sonic section, compatible with additive fuel located on the wall of the combustor.

S Reference cross section of the external front section of the vehicle or engine.

Scramjet Supersonic combustion ramjet.

Solid fuel ramjet (SFRJ) Ramrocket using a solid fuel located on the wall of the combustor.

Specific impulse (I_s) Expressed in seconds, characterizes the consumption (excluding captured air) of an engine for a given thrust ($I_s = F/g \cdot \dot{m}$).

THE RAMJET is an air-breathing propulsion system usable by aircraft during supersonic flight. The scramjet replaces the ramjet above a flight Mach number of around 5 and up to Mach 12 at hypersonic speeds when the air captured from the atmosphere is sufficient.

The main use of the ramjet, now and considered for the future, is the propulsion of supersonic missiles with ranges far beyond those of missiles of an equivalent size using a solid propellant rocket engine as the propulsion system.

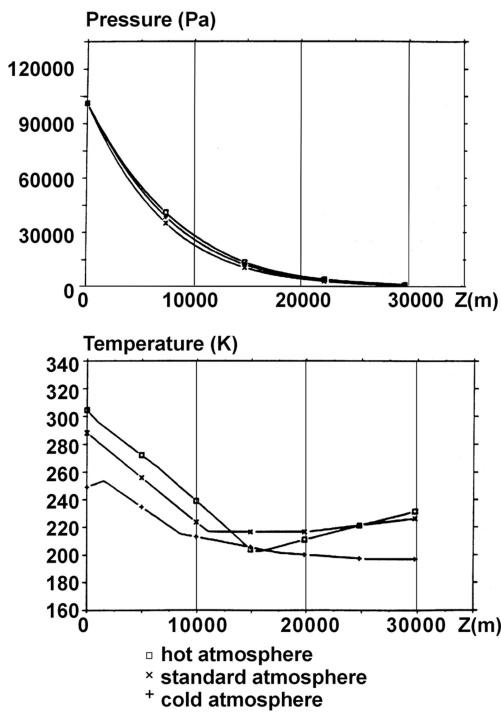


FIGURE 1 Static temperature and pressure of the atmosphere.

The use of the scramjet is being considered for projects for missiles, space shuttles, and hypersonic aircraft, especially because of the low relative fuel consumption and the lower complexity of the structures that have to be capable of withstanding the high temperatures corresponding to flight at high Mach numbers.

I. HISTORY AND POSSIBLE APPLICATIONS

The history of the “aerothermodynamic nozzle”, later called ramjet, is indissociably linked with René Leduc (1898–1968), a Frenchman of modest origins, who filed the first patent application for a pulse jet engine in 1930, followed by a patent application for an aerothermodynamic nozzle in 1933. He then discovered that another Frenchman, René Lorin (1877–1933), had already imagined and described this propulsion principle back in 1913,

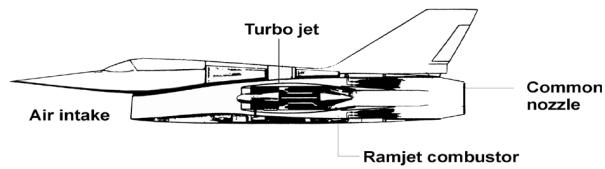


FIGURE 3 Section of the Nord Aviation Griffon 02 aircraft.

without being able to demonstrate any practical advantages for aircraft propulsion, since his experimental devices were too inefficient.

Among those who corresponded with René Lorin and were also interested in jet propulsion were the American Octave Chanute, the Russian Gorokhoff, and the Italian Crocco.

Compared with the turbojet, another form of air-breathing propulsion, the main advantage of the ramjet is the absence of moving parts, leading to a lower cost. One of its major drawbacks is that it is incapable of providing a thrust greater than the drag at low speeds, which means that it requires an initial booster.

From 1946 to 1957, René Leduc built and flight tested several experimental aircraft, all dropped from airplanes. To achieve a positive thrust-to-drag ratio at subsonic speeds, his aircraft featured a minimized drag configuration: The pilot was housed inside the front air intake cone (Fig. 2).

It was Nord Aviation’s experimental Griffon aircraft that made the most striking demonstration of the ramjet from 1957 to 1961 with its combined turboramjet (Fig. 3), which achieved a climb speed of Mach 2.19.

Work on ramjets was stopped in the 1960s when turbojet design was perfected and they were shown to have a lower fuel consumption than ramjets.

At the same time, especially in the United States, Russia, and France, work was being conducted on ramjet use for missiles, where the low cost of the engine is an essential factor, especially when high speeds are required at low and high altitudes, inaccessible to turbojets because of the high temperatures of the air captured at the corresponding flight Mach numbers (Mach 2 to 5).

Experimental missiles (ONERA’s Stataltex which reached Mach 5 at 25,000 m at the end of the 1960s) and the

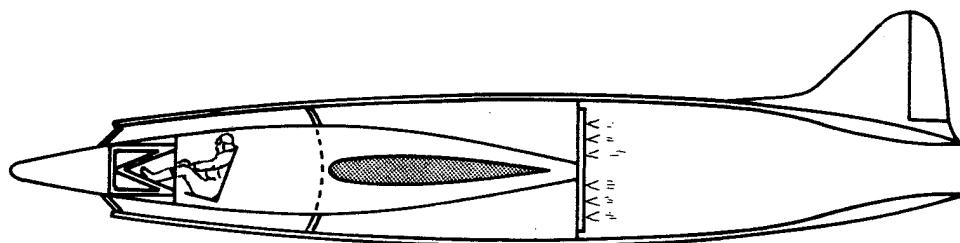


FIGURE 2 Section of an experimental aircraft made by René Leduc.

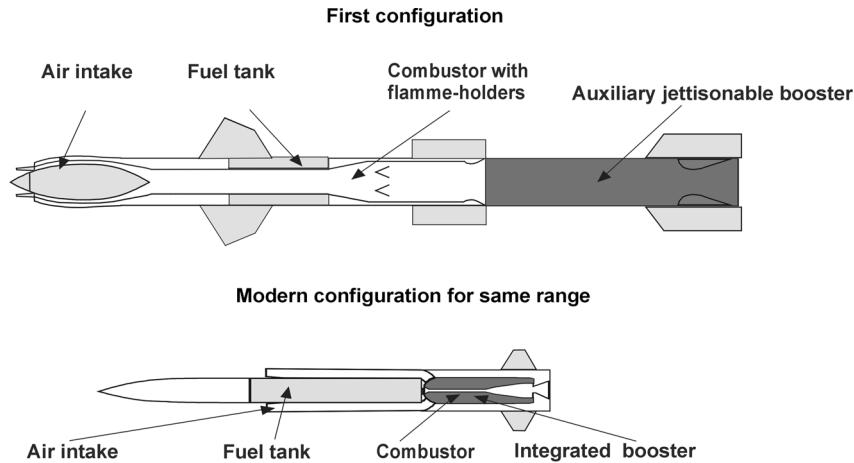


FIGURE 4 (a) Architecture of old generation ramjet missiles. (b) Architecture of new generation ramjet missiles.

first applications, SA4 Ganef (Russia), Navajo, Bomarc and Talos (United States), and Bloodhound and Sea Dart (UK), used a first configuration of the engine which included a conical front air intake, a kerosene fuel tank, a combustion chamber with flameholders, and a tandem booster to accelerate to ramjet speed (after which it was then jettisoned) (Fig. 4a).

It was before the 1970s that the western world discovered the Soviet SA6 Gainful missile with a new configuration. This modern configuration was characterized by one, two, or four lateral air intakes, freeing the entire forward section of the missile for the seeker guidance system and the warhead; a liquid or solid fuelled tank; a combustion chamber without flameholders; and a solid booster engine housed in the combustion chamber, all of whose components disappeared during the boost-cruise transition phase (Fig. 4b).

This configuration resulted in a very compact missile with a long range compared with a solid propellant missile and almost double the speed, even at zero altitude, compared with a subsonic turbojet-propelled missile.

Many demonstrators were developed and tested in the United States (ALVRJ, ASALM, VFDR), Russia, France (Modele Probatoire and MPRS), and Germany (EFA) at the end of the 20th century. So far, only Russia and France (Fig. 5) have adopted this propulsion concept for operational missiles.

Similarly, the missions accessible to small missiles (air-to-air, antiship, air-to-surface, antiradar, surface-to-air) could be performed in the near future by ramjet missiles.

Furthermore, ramjet followed by scramjet propulsion is being considered for hypersonic vehicles in addition to supersonic missiles. Since 1970, there have been projects for space launchers, hypersonic aircraft (U.S. NASP project, Fig. 6), reconnaissance UAVs, and missiles with increased defense penetration capability. This is known as dual

mode ramjet propulsion, subsonic ramjet combustion from Mach 2 to Mach 5, and then changing to scramjet combustion up to Mach 10 or more. Beyond that, rocket propulsion can take over, considering the thermal limitations and the rarefaction in oxygen.

After research and testing of circular section scramjets, such as the HRE in the United States, Esope in France, and Kholod in the former USSR, work in Russia, the United States, Japan, France, and Germany is focused on rectangular-section combined ramjets, not as well adapted to pressure stresses but which fit well under aircraft wings, optimizing engine integration on the aircraft and the thrust-to-drag ratio (Fig. 7).

For instance, for a VTOL (Vertical Take Off and Landing) space shuttle, a rocket engine or turbojet can be used as the initial booster, then relayed by a ramjet, then by a scramjet, and finally by a rocket engine for the exoatmospheric phase. Engines on which these different types of propulsion are closely associated are called combined cycle engines.



FIGURE 5 Aerospatiale ramjet-powered ASMP nuclear missile carried by Mirage 2000, operational in French forces since 1986 (CEV photo).

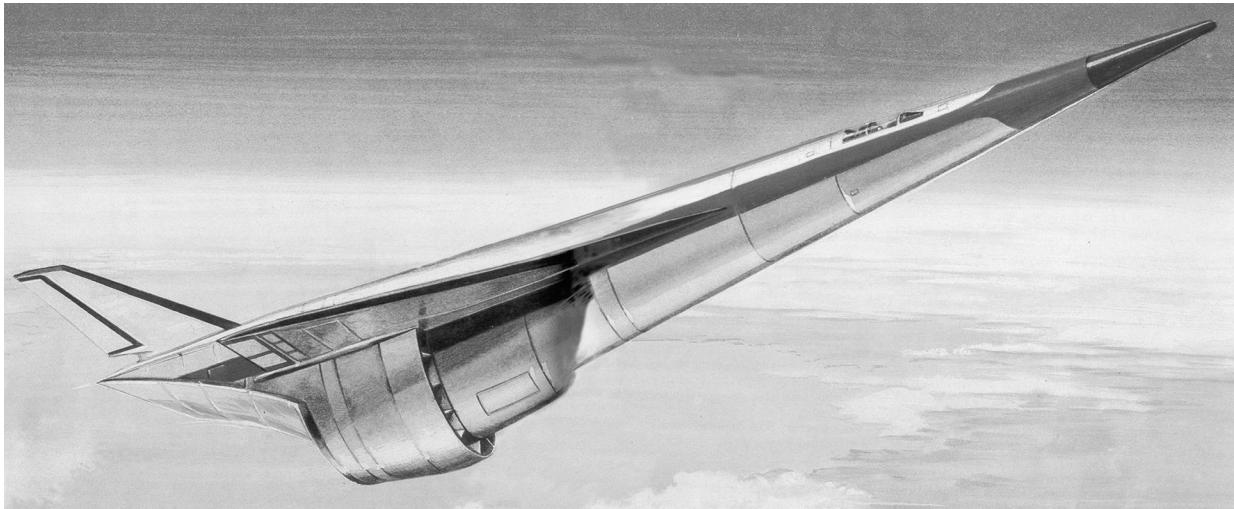


FIGURE 6 American NASP (National Air-Space Plane) project.

II. RAMJET OPERATION

A. Performances

Although variable geometries may be considered, in particular for the air intake or nozzle section, to optimize operation in all points of flight, the main advantage of the ramjet is the lack of moving parts. This compares favorably with the turbojet, other air-breathing concept, which includes sophisticated parts rotating at very high speed. Actually, when comparing the specific impulse (representative of fuel consumption), flight Mach number, and flight altitude, it can be seen that each type of propulsion is suited to a particular range (Fig. 8).

Liquid or solid rocket propulsion, which does not require the presence of air and can be used at any Mach number and any altitude, satisfies the need for strong, short-duration thrust. In the atmosphere, cruise propulsion is efficient by turbojet at low altitude and low speed and then by ramjet and scramjet at high speed and low or high altitude.

B. Atmosphere, Mach

Like turbojets, ramjets and scramjets take the oxygen required for combustion from the air during flight, contrary to rockets, which use liquid, solid, or even gaseous fuel and oxidizer.

Vehicles equipped with air-breathing engines therefore carry only fuel in their tank. This results in a low combustion of on-board products, or a high specific impulse. However, the capture of air during flight has a cost, requiring air intakes whose weight and size are not negligible. In addition, the captured flow varies according to many parameters such as the Mach number and altitude, but also the vehicle attitude (angle of attack, yaw, roll angle).

As can be seen in Fig. 1, the atmosphere varies enormously according to the altitude. The static pressure is halved on the average every 5000 m. The local static temperature, equal to 278.15 K in a standard atmosphere, drops to 216.65 K at 11,000 m and remains constant up to 20,000 m. The pressure and temperature also vary according to the type of climate.

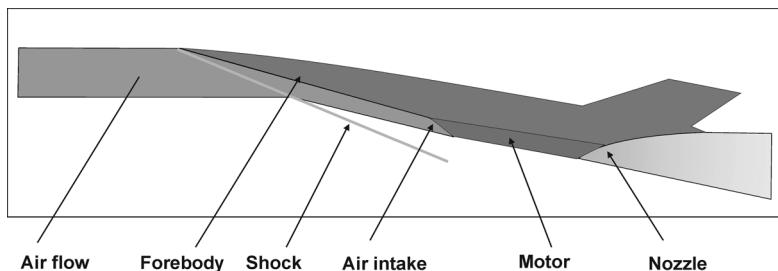


FIGURE 7 Airframe-ramjet integration.

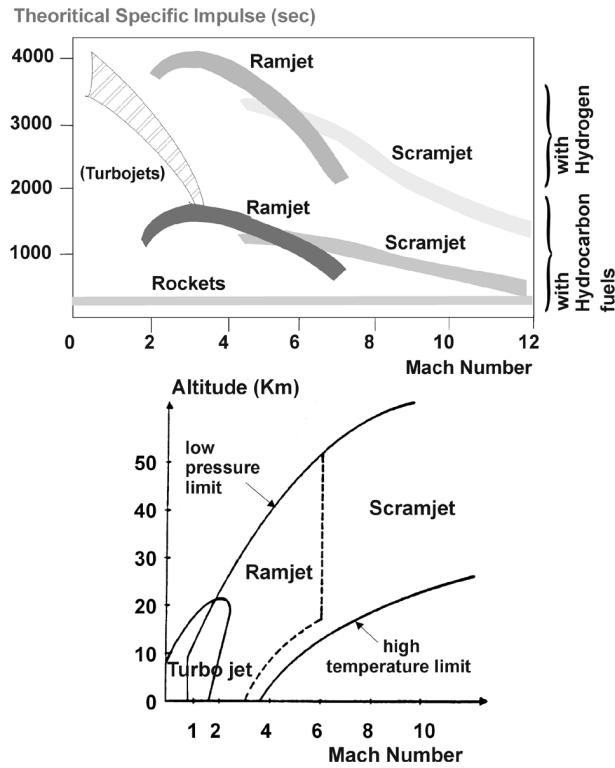


FIGURE 8 Ranges of use and theoretical performance of various propulsion modes.

In addition to the static pressure and temperature, the total pressure and temperature are parameters which characterize the stages of the propulsion system. The following equations are used to convert easily from static values to total values:

$$T_T = T \cdot \left(1 + \frac{\gamma - 1}{2} \cdot M^2\right),$$

$$p_T = p \cdot \left(1 + \frac{\gamma - 1}{2} \cdot M^2\right)^{\frac{1}{\gamma-1}}.$$

Speed is also one of the dimensioning characteristics of the flow captured by the air intakes. Since the flight conditions are generally supersonic, the local flight Mach number or the Mach number inside the ramjet is generally used instead of the speed.

C. Ramjet Components

1. Diagram of a Ramjet

[Figure 9](#) shows the subscripts normally used for the different characteristic sections of the flow of air or burnt gases, from free stream (subscript 0) to the nozzle exit (subscript 6).

2. Air Intakes and Diffuser

There are many types of air intakes. Their geometries generally correspond to integration on the fuselage according to it they are frontal (Pitot, conical, isentropic, or chin air intakes) or lateral (conical, semicircular, two-dimensional air intakes). [Figure 10](#) shows a two-dimensional lateral air intake. For each vehicle, the configuration selected results from a multitude of constraints.

The captured air flow is expressed as follows:

$$\dot{m}_0 = \frac{\gamma_0 \cdot p_0 \cdot M_0 \cdot A_1 \cdot \varepsilon}{a_0}.$$

The relative weight flow of the air intake (ε) depends on flight conditions such as the angle of attack or yaw and on its mode of operation.

The shock on lip Mach number (M_{sol}) of an air intake corresponds to the fact that the first oblique shock from the flat or circular ramp has just touched the opposite lip of the air intake. Below this Mach number, the relative weight flow is less than 1 (at zero angle of attack and yaw).

[Figure 11](#) shows these different types of operation: supercritical, critical, or subcritical for a given flight Mach number, angle of attack, yaw, and roll angle. The inlet

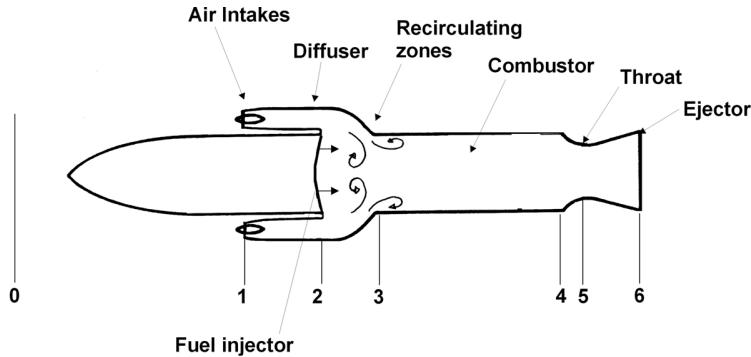


FIGURE 9 Parts and reference sections of a ramjet.

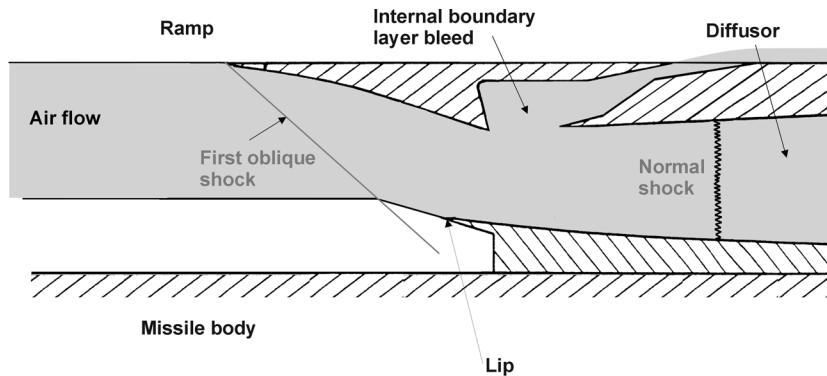


FIGURE 10 Example of a two-dimensional Mach 2 to 4 air intake.

pressure recovery (η_{02}) is the ratio between the total pressure at the end of the diffuser and the free-stream total pressure. For a given angle of attack, yaw, and roll, the critical inlet pressure recovery and critical relative weight flow of an air intake with a given geometry vary as shown in Fig. 12.

The higher the inlet pressure recovery is, the more optimized the ramjet operation, especially in terms of specific impulse.

The air intake is actually designed to convert a supersonic free-stream flow to a subsonic flow at the combustion chamber inlet. This is accompanied by an increase in the static pressure. The transition from a supersonic flow to a subsonic flow takes place in the diffuser, with an increasing section, through a normal shock. When the pressure in the combustion chamber is low (corresponding to a low fuel mass flow rate), the normal shock is relatively far from the inlet of the air intake. If the combustor pressure is too high, the shock moves upstream toward the air intake and there is a risk of buzz. This phenomenon is to be avoided because it can lead to structural damage or a deviation in

the vehicle trajectory, since the burning flames come out of the air intakes.

The numerical models now available are capable of calculating the characteristics of an air intake for simple cases of angle of attack and roll angle, but an experimental approach is still required to obtain all the characteristics for all values of Mach number, angle of attack, yaw, and roll angle. A complete model on a significant scale of the front part of the vehicle extending one or two diameters aft of the air intake is placed in a wind tunnel suited to the flight conditions to be reproduced. This model, fitted with a large number of pressure sensors required to obtain the flow profiles, is equipped aft of the air inlet into the combustor with a variable section shutter, allowing the pressure to be varied in Section 3 in order to measure the operating conditions of the air intake from supercritical to buzz as the pressure increases.

3. Fuels

Kerosene is the most commonly used fuel for ramjets because of its high energy, the ease of mass flow rate control, its suitability for storage, and its low cost. Other liquid fuels, often costing more, can be used especially when a higher density is required to increase the ramjet operating time for a given fuel tank size. There are also both liquid and slurry fuels containing a high proportion of solid boron particles whose energy-to-volume ratio is especially advantageous. Nevertheless, the use of such fuels involves other difficulties, in particular for the fuel mass flow rate control.

The best fuel is hydrogen because it ignites rapidly in air and yields high specific impulses. However, because of its very low density compared to liquid and solid fuels, excessively large fuel tanks are required. This leads to the use of another fuel property: the volumetric impulse ($\rho \cdot I_s$).

Table I gives a few characteristic values of the different fuels. The stoichiometric ratio of a fuel (f_s) is a ratio

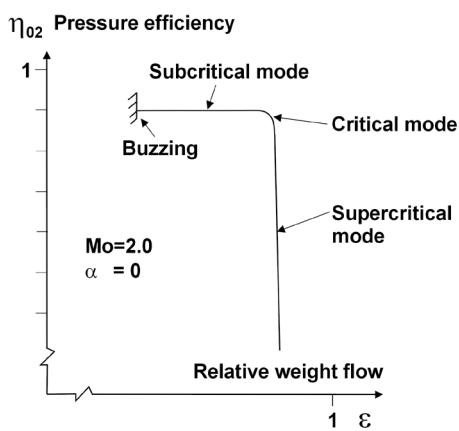


FIGURE 11 Characteristics of an air intake at a given Mach number.

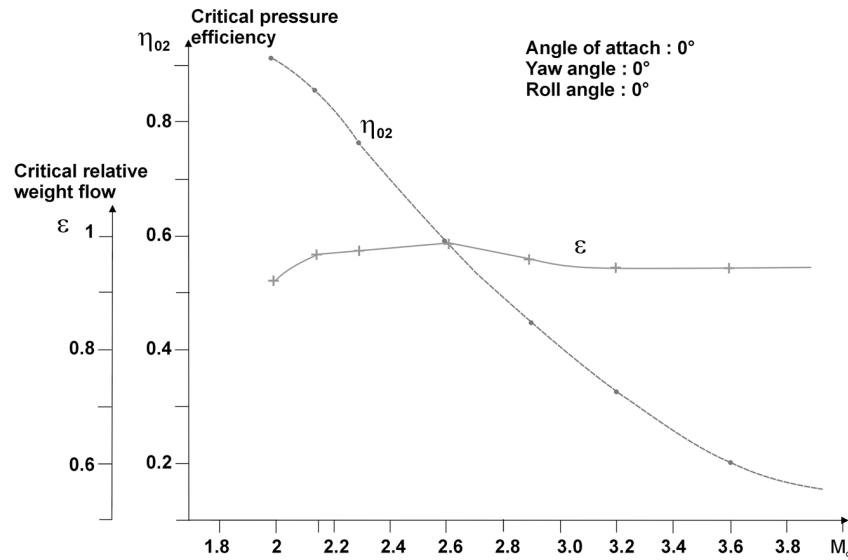


FIGURE 12 Variation of air intake critical point coordinates versus Mach number.

between the fuel mass flow rate and the air mass flow rate, leading to complete combustion of the components, i.e., a fuel-air (F/A) equivalence ratio (φ) equal to 1.

The value of the stoichiometric ratio depends on the composition of the air and, in particular, the oxygen content, often assumed equal to 23.48%. However, the air composition is not a universally used value.

Solid fuels may have the form of a very fine powder (e.g., aluminum) or a solid fuel grain, similar to the solid propellants used for rocket propulsion. A solid propellant is a mixture of oxidizer (such as ammonium perchlorate) and binder (such as polybutadiene) in optimized proportions to achieve complete combustion at high energy. A basic solid fuel uses the same components, but with only the percentage of oxidizer necessary for its combustion with a quantity of binder to transform the rest of the binder to a combustible gas which burns with air in the ramjet combustor. This type of combustion is known as pyrolysis.

Solid fuel formulation specialists have to find a trade-off among the ingredients and operating modes between the possibility of manufacturing the grains, the mechanical properties, the energetic performance (Is), the density,

the burning rate, the sensitivity to operating pressure and temperature, etc.

4. Combustor

A modern combustor is supplied with air captured by the air intakes and with fuel through optimally located injectors, without flameholders, therefore making the best use of the flow recirculation possibilities to obtain the best combustion efficiency, lowest total pressure loss, and the F/A equivalence ratio operating range suited to the mission of the vehicle.

The method described herein to define the performance of a combustor uses the concept of characteristic velocity (C^*) of the air-fuel combustion. The characteristic velocity $C^*(\varphi_b, h_{T_0})$ (in m/s) as a function of the air total enthalpy [$h_{T_0} = T_0 \cdot (1.006 + \gamma_0 \cdot R \cdot M_0^2 / 2000)$ in kcal/kg] and the F/A equivalence ratio (φ_b) are calculated from JANAF tables. These calculations also give the value of $\gamma(\varphi_b, h_{T_0})$ at the ramjet throat. **Table II** gives the characteristic velocity and γ of the kerosene as a function of the burned F/A equivalence ratio for two total

TABLE I Some Properties of Ramjet Fuels

Characteristics	Liquid fuels			Solid fuels		
	Kerosene	High density	Boron slurry	Ducted rocket	Rustique ducted rocket	With boron
Density	0.786	0.939	1.307	1.189	1.259	2.025
f_s	0.0685	0.0713	0.0868	0.117	0.156	0.235
$\rho \cdot I_{s\text{theor.}}$ at $M_0 = 2$	1302	1515	2389	1289	1182	2053
$C^*(\varphi_b = 0.5, h_{T_0} = 520)$	1046	1049	1117	1062	1102	1183
$\gamma_5(\varphi_b = 0.5, h_{T_0} = 520)$	1.287	1.286	1.257	1.283	1.270	1.250

TABLE II Characteristic Velocity and γ of Kerosene

Equivalence ratio	0.2	0.4	0.6	0.8	1.0	1.2
$C^*(h_{T0} = 520)$	811	978	1105	1206	1274	1273
$C^*(h_{T0} = 800)$	901	1047	1161	1248	1301	1312
$\gamma_5(h_{T0} = 520)$	1.334	1.299	1.277	1.262	1.254	1.259
$\gamma_5(h_{T0} = 800)$	1.319	1.291	1.272	1.260	1.255	1.258

enthalpies. The variations are practically linear between these two enthalpies.

The characteristics used to reproduce the combustor operating condition are as follows:

- The injected equivalence ratio ($\phi_i = \dot{m}_f / (\dot{m}_0 \cdot f_s)$)
- The combustion efficiency ($\eta_c = \phi_i / \phi_b$) where ϕ_b is the burned F/A equivalence ratio corresponding to the injected fuel flow rate required to obtain the same combustion performance, when the efficiency is equal to 1
- The total pressure recovery ($\eta_{2-5} = p_{T_5} / p_{T_2}$)
- The combustor burned F/A equivalence ratio operating range between the lean burnout ratio and the rich burnout ratio; in practice, it is not always necessary to be able to operate at burned F/A equivalence ratios above 1, but combustion up to ratios above 3 can, in some cases, avoid engine burnout

Specialists developing such combustors attempt to optimize all these parameters using numerical models and test facilities (Section VII).

Figures 13 give a realistic representation of the variations in combustion efficiency and pressure recovery of a combustor.

Stabilization of the combustion is generally realized by air and fuel recirculation regions and optimized combustion regions at various F/A equivalence ratios. This leads to local temperature gradients which may cause temperature rises above those the thermal insulation on the wall of the combustor is capable of withstanding.

There is also the risk of occurrence of instabilities in certain operating ranges. Specialists use their experience and know-how to solve these problems.

Generally, the air-fuel mixture in the combustor does not ignite spontaneously. This is the case in particular for the air-kerosene mixture. The use of a suitably dimensioned and located igniter is then necessary to initiate combustion, which must be sustained throughout the duration of the flight.

5. Nozzle

In a ramjet, the combustion gas pressure and temperature are converted to thrust through the nozzle, including a

convergent, a sonic throat, and a divergent. Because of the presence of a boundary layer, the sonic section is reduced by a flow coefficient (C_{d5}) of around 0.99 compared to the geometric section of the throat.

In addition, nozzle efficiency (η_j) is used to take into account the fact that expansion in the nozzle is not isentropic and, because of the approximation used to take γ_5 equal to γ_6 . It generally has a value of around 0.985.

For the thrust, it is advantageous to attempt to increase the sonic section of the nozzle, but this can lead to difficulties in stabilizing the combustion. So, the Mach number in the combustor is often limited to less than 0.5.

The exit section is optimized as a function of the diverging angle and the length of the nozzle. The nozzle exit section is never greater than the front section of the vehicle, in particular to limit drag.

D. Ramjet Thrust

1. Thrust Calculation

Knowing the engine geometry, the flight conditions, the mass flow rate of the air entering the ramjet (which

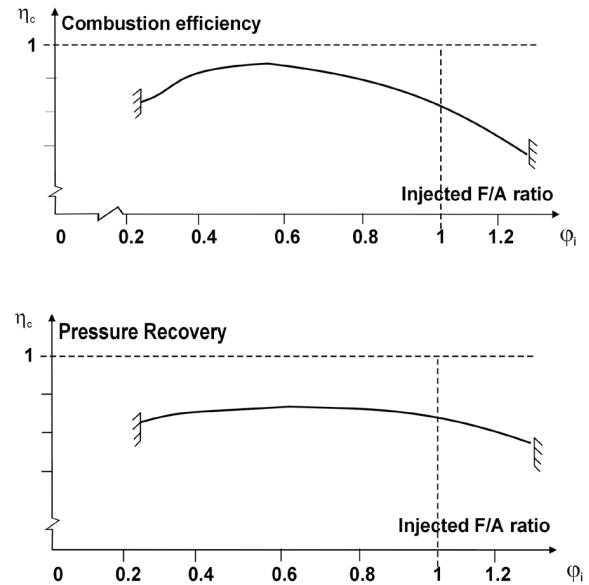


FIGURE 13 Example of ramjet combustion efficiency and pressure recovery variation.

depends on the precision of the relative weight flow ε , or on the air intake efficiency η_{02}), and the injected mass flow rate of fuel \dot{m}_f , the injected equivalence ratio φ_i can be determined. From this, can be known the combustor pressure efficiency η_{24} , for the engine and fuel considered, as well as the combustion efficiency η_c and thus the burned equivalence ratio φ_b , the characteristic velocity C^* , and γ_5 . The mass flow rate at the nozzle sonic throat is therefore

$$\dot{m}_5 = \frac{A_5 \cdot C_{d_5} \cdot p_{T_5}}{C^*} = \dot{m}_0 + \dot{m}_f,$$

where the stagnation pressure at the throat is

$$p_{T_5} = \frac{(\dot{m}_0 + \dot{m}_f)}{A_5 \cdot C_{d_5}} \cdot C^*$$

and the required air intake efficiency, which should correspond to the value initially taken for its relative weight flow, is

$$\eta_{0-2} = \frac{p_{T_5}}{p_{T_0}} \cdot \frac{1}{\eta_{2-5}}$$

with $\eta_{0-2} = \eta_{0-2}(M_0, \alpha, \dots)$.

When the nozzle is completely energized, we can deduce from the conditions in 5 the nozzle exit conditions 6, notably, p_6 , M_6 , and the exit dynalpy D_6 ,

$$M_6 = M_6(A_5/A_6, C_{d_5}, \gamma_5, \dots)$$

$$D_6 = p_6 \cdot A_6 \cdot (1 + \gamma_6 \cdot M_6^2)$$

or

$$D_6 = \eta_j \cdot A_6 \cdot \frac{p_{T_5}}{\left(1 + \frac{\gamma_6 - 1}{2} \cdot M_6^2\right)^{\frac{\gamma_6}{\gamma_6 - 1}}} \cdot (1 + \gamma_6 \cdot M_6^2)$$

For any air-breathing engine, the thrust is expressed as a function of the exit dynalpy D_6 of the exit streamline, the entrance dynalpy D_0 of the entering streamline, and the pressure correction on the entry and exit sections.

So, for a vehicle with incidence and with no yaw, the thrust along its x axis is

$$F_x = D_6 - D_0 \cdot \cos \alpha - p_0 \cdot (A_6 - A_0 \cdot \cos \alpha).$$

A thrust coefficient C_F can be deduced from this, which is comparable to the drag coefficient C_x :

$$C_F = \frac{F_x}{\frac{1}{2} \gamma_0 \cdot p_0 \cdot M_0^2 \cdot S},$$

and the specific impulse is

$$I_s = \frac{F_x}{g \cdot \dot{m}_f}.$$

Realistic values. For a ramjet missile using kerosene at an equivalence ratio of 0.5 and flying at Mach 2 at sea level, the following characteristics can be obtained:

$$\begin{aligned} p_{T_0} &= 792,812 \text{ Pa} \\ t_{T_0} &= 520 \text{ K} \\ p_{T_2} &\approx 700,000 \text{ Pa} \\ M_2 &\approx 0.36 \\ M_4 &\approx 0.44 \\ p_{T_5} &\approx 600,000 \text{ Pa} \\ t_{T_5} &\approx 1640 \text{ K} \\ M_6 &\approx 1.9 \\ p_6 &\approx 900,000 \text{ Pa} \\ I_s &\approx 1450 \text{ s} \end{aligned}$$

2. Thrust Limitation

When the fuel mass flow rate becomes too high,

- The pressure in the combustor no longer increases strongly as soon as the burnt equivalence ratio exceeds 1.
- The combustor can flame out.
- The air intake can switch to a subsonic regime or even start buzzing.

3. Thrust-Drag Balance

Ramjet thrust depends very much on the static pressure p_0 of the air at the flight altitude. At a given equivalence ratio, it is nearly proportional to it, and the thrust coefficient C_F becomes almost independent of it.

This makes the thrust and drag coefficients comparable at zero or low incidence. [Figure 14](#) shows the different regions in which the ramjet is used:

- Below a given flight Mach number that depends on the engine and air intake geometry, the drag is always greater than the thrust and the vehicle can only decelerate.

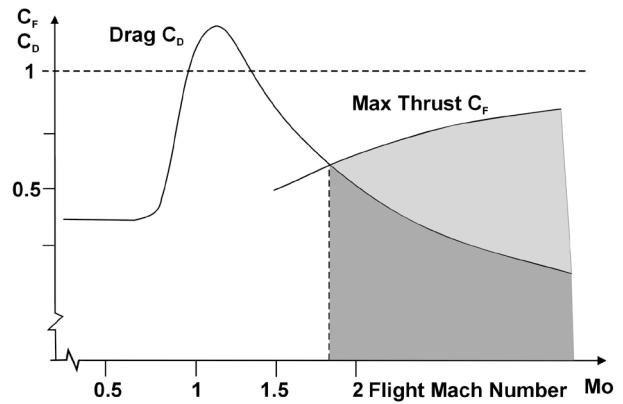


FIGURE 14 Comparison of typical missile thrust and drag coefficients.

- Above this Mach number, the difference between the maximum thrust and the drag is positive, and therefore the vehicle's acceleration capacity increases with the Mach number.

The acceleration capacity is adjusted by adjusting the fuel mass flow rate and thus the equivalence ratio. In cruising flight at constant altitude, the thrust is adjusted to be equal to the drag.

Except for high-velocity projectiles with relatively low mass, the acceleration capacity of a ramjet-propelled vehicle rarely exceeds 1 g.

III. INTEGRATED BOOSTER

In order to reach the flight Mach number needed to achieve sufficient thrust to allow acceleration on their trajectory, modern ramjet-powered missiles use a solid propellant booster integrated into the combustor (Fig. 15). A circular section of the missile and combustor is well suited to this, because the propellant grain operating pressure is about 10 MPa.

In order to seal the combustor during the acceleration phase, the air intakes into the combustor have to be blanked with port covers, which are later opened and often ejected in the transition phase between acceleration and cruise.

To optimize the solid propellant booster performance and achieve a specific impulse of the order of 240 s throughout operation, a specific nozzle should be fitted to the exit, generally with a section less than that of the ramjet, so it can advantageously be included inside the ramjet nozzle. This nozzle also has to be ejected during the transition.

The solid propellant grain is generally cast directly on the thermal insulation inside the ramjet combustor. The shape of its burning area should be defined to optimize the filling coefficient and to reach the end of burn as quickly as possible. This minimizes the boost-cruise transition duration without reducing the vehicle's velocity and releases all the fuel from the ramjet combustor to avoid any risk of air intake buzzing when the port covers open and the air starts to enter.

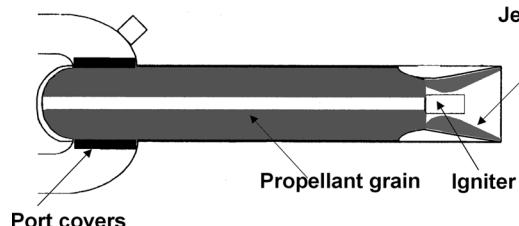


FIGURE 15 Integrated booster with jettisonable nozzle.

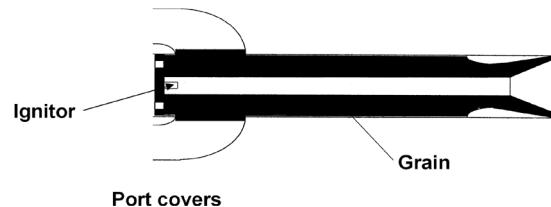


FIGURE 16 Integrated nozzleless booster.

The transition phase duration is often defined as the phase between the ejection of the first port covers (if the engine has several air intakes) and the moment when the ramjet reaches 90% of its nominal thrust. Transition durations of between 0.2 and 0.5 s are achievable.

The ramjet combustor length is thus not generally optimized in terms of the combustion kinetics between fuel and air, but rather for the booster volume needed to increase the flight Mach from Mach 0 (or about Mach 0.6 for a missile fired from a combat aircraft) to the transition Mach number.

Another type of integrated booster is advantageous for use on missiles of limited diameter: the integrated nozzleless booster. As shown in Fig. 16, this is a solid propellant grain with circular bore of optimized section, ending with a nozzle divergent formed directly by the grain itself. In the course of operation, the local burning rate varies the shape of all of the sections, including that of the nozzle. The reduced specific impulse of 200 to 220 s can be sufficient, considering the additional volume of propellant and the weight saved on the metal nozzle. Furthermore, the very fact of not having to eject a metal nozzle is a sure safety advantage, especially for missiles fired from aircraft.

IV. LIQUID FUELLED RAMJET

Figure 17 shows a diagram of a liquid fuelled ramjet. The fuel is usually kerosene.

In storage, aside from the solidification or self-ignition limits to be considered, the fuel tank is generally kept at a slight pressure by an expendable device to compensate the thermal expansions of the liquid and avoid sloshing, especially when being carried by an aircraft.

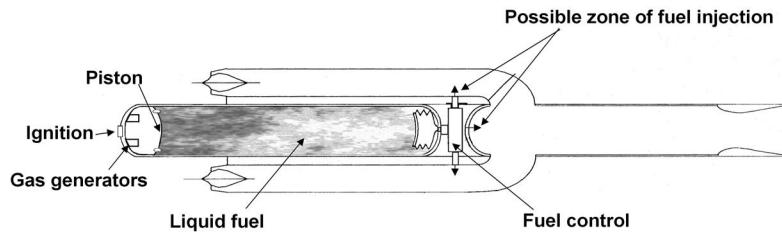


FIGURE 17 Diagram of a liquid fuel ramjet.

In order to ensure a long life without alteration of the fuel and the tank components (especially the seals), the liquid is generally subjected to special treatments. A blanking plug is ruptured at the time the engine starts service to break the tank seal.

The fuel mass flow rate is generated and regulated by tank pressurization, with a piston-opening system using solid propellant gas generators that are ignited as the piston advances and the free volume increases, or by a turbine pump driven by pressurized air taken from the air intake.

Regulating a mass flow rate of liquid generally poses no problems. The difficulties to be solved have to do with the required precision in the regulation range between maximum and minimum flow rates. Regulation rates in the vicinity of 25 are quite easily possible, with very short response times.

Once the fuel is injected in the combustor, one of the problems to be solved is to make sure it burns with the air. Injectors are commonly used to vaporize the liquid in the form of fine droplets by a vortex effect.

Combustion with the air is generally not initiated spontaneously. A high-energy igniter is needed, and sometimes two are used for reliability.

The liquid fuel ramjet operating sequence with its integrated booster is thus generally the following:

- Booster ignition and burn (2–5 s)
- Opening of the air intakes (especially for airborne missiles)
- Tank pressurization
- Detection of end of booster thrust
- Nozzle cut and ejection
- Ejection of port covers and admission of air
- Admission of fuel in the combustor

- Combustor ignition
- Regulation of fuel mass flow rate in accordance with trajectory and flight instructions

This means a minimum of six pyrotechnic orders or commands must be sequenced in a limited time, or four such orders for a nozzleless booster.

V. SOLID FUELLED RAMJET

A. Solid Fuel Ramjet (SFRJ)

The ramjet powered by solid fuel bounded on the wall of the combustor is an American specialty, especially of the CSD company. As Fig. 18 shows, the air is inlet into the combustor along the centerline, with the possibility of regulating the air mass flow rate and thus the thrust.

Beyond the region including the fuel, there is a second lower Mach combustor that can be supplied with air from a second air intake. This advantageously increases the combustion efficiency, and an integrated booster can be lodged in this section.

Very low burning rates and very high burning efficiencies are possible, including with very energetic fuels containing boron.

This configuration is well suited to artillery projectiles, for which no integrated booster is needed. For other purposes, it seems difficult to achieve high filling coefficients for the booster.

B. Ducted Rocket

The United States (ARC), the former Soviet Union, Germany (Bayern Chemie), and France have worked on this ramrocket concept.

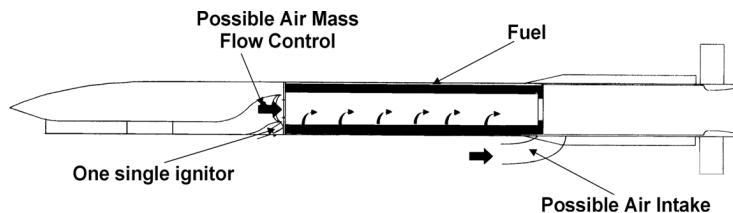


FIGURE 18 Diagram of a solid fuel ramjet.

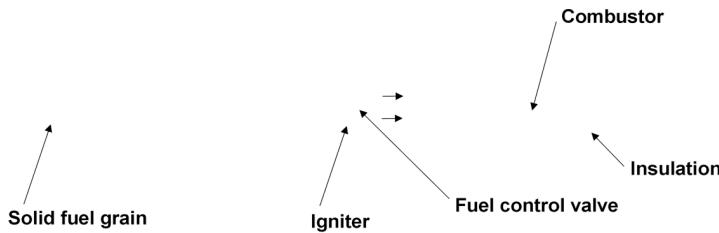


FIGURE 19 Diagram of a fuel controlled ducted rocket.

Solid fuel gas generators use grains similar to those of solid propellants. Their burning area is generally constant, either with radial burning or front-face burning as shown in Fig. 19. The burning rate and combustion area have to be designed to generate the desired gas mass flow rate.

The burning rate generally ranges from 0.2 mm/s for radial burning at low pressure to 25 mm/s at high pressure for front burning. Once ignited by an appropriate initiator, the burning area A_b emits a burned fuel mass flow rate \dot{m}_b , which is expressed in the form

$$\dot{m}_b = A_b \cdot v_b \cdot \rho_f,$$

where ρ_f is the density of the fuel and v_b is the burning rate, which is sensitive to the combustion pressure p_g in the generator and to the initial grain temperature θ_i compared to a reference temperature θ_r , according to the formula

$$v_b = a \cdot p_g^n \cdot e^{(\theta_i - \theta_r)}.$$

The pressure is controlled by the sonic section A_s of a valve at the generator exit, and the mass flow rate \dot{m}_f of the gaseous fuel injected into the combustor is equal to

$$\dot{m}_f = \frac{A_s \cdot p_g}{C^*}.$$

In steady regime, the burned mass flow rate of the gas generator and of the throat exit are equal. As Fig. 20 shows, it is then possible to adjust the mass flow rate between a minimum and maximum, regardless of the grain temperature, for a given constant burning area. To increase

the mass flow rate modulation rate, the fuel's burning rate should have a high-pressure exponent n (but not greater than unity) with low-temperature sensitivity. The sensitivity a is generally in the neighborhood of 0.3% per Celsius degree, but can range from less than 0.1 to 1% (which is prohibitive). So, for a reasonable generator operating range between 1 and 150 bar, a maximum modulation rate of 10 is possible but difficult to achieve. It should be noted that, to remain sonic, the pressure in the gas generator should always stay twice as high as the pressure in the ducted rocket combustor.

For good flow rate precision, it is important to minimize the dispersions in the fuel's burning rate and to have a combustion area that remains always constant.

At maximum pressure, for a gas generator of a 180-mm caliber missile, the equivalent diameter of the sonic throat is often only 2–4 mm, which is extremely delicate for the regulation device to regulate, generating a major risk (of explosion) for the generator case if any solid deposits crystallize on this sensitive section.

The regulation valve is generally lodged between the gas generator and the combustor, where the gases are injected at the level of the front. Once initiated (usually spontaneously, without specific igniter), the air and fuel generally burn with no problem, considering the ease with which these two gaseous components can mix, unless the fuel includes metallic or metalloid particles leading to the presence of solid phase oxides at the generator outlet. This is especially true of fuels with boron additive to increase the available energy.

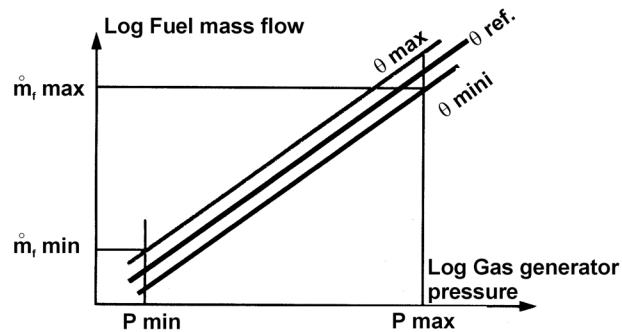


FIGURE 20 Flow rate control of a gas generator for a ducted rocket.

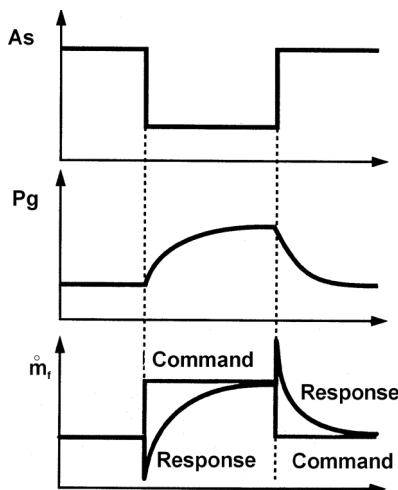


FIGURE 21 Pressure and fuel flow rate response of a controlled gas generator for a ducted rocket.

In light of the gas mass variation in the open volume of the gas generator in transition regime, the exiting mass flow rate is expressed as follows:

$$\dot{m}_f = \dot{m}_b - \frac{dM}{dt}.$$

As Fig. 21 shows, in order to increase the mass flow rate injected, reducing the throat section gradually increases the pressure for a few seconds (which depends on the open volume, i.e., if much or little solid fuel is remaining). However, the resulting mass flow rate passes through an under-flow phase before reaching the required value. In the reverse direction, an excess mass flow rate is obtained. These phenomena, which can lead to engine buzzing or flameout, are harmful for thrust regulation. A lesser evil would be a delay between the command and response, when the necessary flow rate variation is slow. This is a huge disadvantage compared with flow regulation for a liquid fuel.

The operating sequence for a ducted rocket with a controlled gas generator and integrated nozzleless booster is generally as follows:

- Booster ignition and burn (2–5 s)
- Opening of air intakes (especially for airborne missiles)

- Detection of booster thrust decay
- Ignition of gas generator
- Breakage of sealing plug, and admission of gases into the combustor
- Port covers ejection and air admission
- Combustor ignition
- Fuel mass flow rate regulation in accordance with trajectory and flight instructions

C. Rustique Ducted Rocket

The rustique ducted rocket concept was invented in France, at Onera, after work had been done on nearly all the concepts and on quantifying their advantages and drawbacks. Work on this concept, starting in 1978, was oriented toward ramrocket applications to mass-produced low-caliber missiles with storability, maintenance, and reliability characteristics that were to be comparable with those of solid propellant missiles. This means using solid fuel, looking for simplicity, especially by suppressing the active flow regulation, and minimizing costs, even if it means less performance. Why try for a specific impulse six times greater than that of a solid propellant rocket if, with a multiplier coefficient of only 3 or 4, we can get a missile whose range, velocity, or maneuverability is three times greater than that of the rocket-propelled missile of the same size?

All of these elements led to the definition of the “rustique” concept presented in Fig. 22. This concept includes

- A nozzleless booster
- A gas generator lodged in the same case and separated from the combustor by nonsonic sections; the purpose of these sections is just to inject fuel gases at the best points on the combustor front and get the best combustion efficiency
- A single igniter to initiate both the booster and the gas generator
- Air intakes, preferably four of these, two-dimensional inversed, in order to be able to perform high-incidence skid-to-turn maneuvers in all directions
- A solid fuel, whose formulation allows spontaneous ignition of the combustor, while the burning rate with a high pressure exponent (up to 2.5) is suited to the

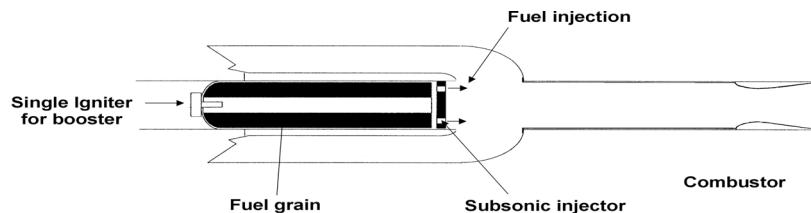


FIGURE 22 Diagram of a rustique ducted rocket.



FIGURE 23 View of an MPSR demo missile with a rustique ducted rocket.

necessary mass flow rate and to the combustion area, in the nominal operating range between 10 bar (at low altitude and high velocity) and 0.3 bar at high altitude

One of the advantages of a nonsonic gas generator is its capacity to adapt the gas generator mass flow rate automatically to the flight conditions. When the missile is flying at low altitude, pressure in the combustor is high and that of the gas generator is also high, as is the mass flow rate of the gas. At high altitude, the mass flow rate of air captured is low, as are the internal pressure and the gas mass flow rate.

This automatic mass flow rate adjustment must be such that the missile is always accelerating when it is at low incidence, whatever the flight conditions. The adjustment parameters are the burning rate and the gas generator surface area (which cannot be constant).

To compute the thrust of a ramjet like this, the generator mass flow rate has to be calculated assuming that the pressure p_g is about 1.1 times the pressure p_{T2} . It should also be noted that, with some of the fuels used in the rustique mode, a combustor can operate at an injected equivalence ratio greater than 3. This is another factor in the automatic thrust regulation process.

The typical rustique ducted rocket operating sequence is as follows:

- Booster and gas generator ignition and burn (2–3 s)
- Spontaneous opening of the air intakes (mainly for airborne missiles)

- Automatic ejection of port covers (by pressure differential) and admission of air
- Continued admission of fuel in the combustor
- Spontaneous combustor ignition
- Fuel flow rate automatically regulated according to trajectory

Figure 23 is a view of one of the seven French MPSR demo missiles.

VI. SCRAMJET

A. Supersonic Combustion

Ramjet performance is directly related to the air intake efficiency but, as Fig. 12 shows, this drops enormously at high Mach numbers. It would be higher if the Mach number were not subsonic at the level of the diffuser, but supersonic; whence the idea of no longer designing the combustor for subsonic combustion as in a ramjet, but for supersonic combustion, making a scramjet. The challenge here is not easy because experience has shown that few ramjet combustors operate beyond Mach 0.5.

The fuel is very important in this case. Hydrogen is well known for its very short inflammation delay and is thus very favorable, especially when it is remembered that the total temperature of air is high at high Mach numbers, which is another favorable factor.

It is true that only scramjets operate to satisfaction with hydrogen today, unless very long combustors are used.

B. Scramjet Components

1. Air Intake

For scramjet applications currently being considered, there are no plans to use boosters integrated in the combustor or to have internal pressures that would be prohibitive for the use of rectangular-section structures. What is usually proposed is rectangular-section air intakes, which make it possible to design variable-geometry air intakes to control the captured mass flow rate either by varying the geometric section A_1 or by varying the relative weight flow ε .

The vehicles used should have a high lift-to-drag ratio for cruising at high Mach numbers; so vehicle sections are usually rectangular also, lenticular or flattened, but never circular. Flat vehicle undersides, or a wing, are then helpful in integrating the air intake under the vehicle. The air intake can be placed in such a way that, when a given angle of attack is needed, it is then possible to use the resulting oblique shock to reduce the entry Mach number in the air intake greatly, to the benefit of its pressure efficiency.

2. Combustor

The supersonic combustors considered have only one axial air inlet.

In a combustor subject to flows at high Mach numbers, it is difficult to cover a large section with fuel injected only from the rectangular or square wall.

If the section is very much flattened, wall injection becomes possible without material obstacle. For larger sections, especially when they are nearly square, injection struts have to be used to cover the entire flow. In either case, it is important that the strut or injected fuel not create a local shock, in which case the flow would become subsonic and induce a large loss in the total pressure.

Injection can be made either perpendicular or axial using obstacles that produce local separations favoring combustion initiation. These axial injectors generally require cooling devices.

Such combustor design requires the definition of zones and a type of injection that correspond as well as possible to the various flight Mach numbers and mass flow rates.

3. Dual Mode Ramjet

A dual mode ramjet is one in which the combustion is subsonic when the flight Mach number is low and becomes supersonic as soon as possible as the flight Mach number increases. This usually means using fuel injection points and modes that vary depending on the conditions and variable-geometry elements from the air intakes to the exhaust nozzle.

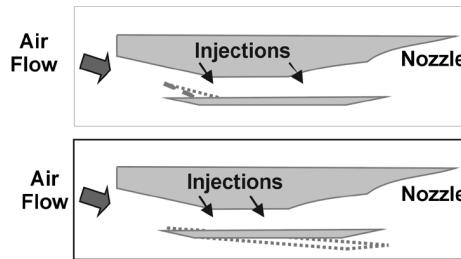


FIGURE 24 Diagram of a scramjet.

While it may seem easy to use variable geometries on paper, constructing a system without leakage is not so easy, especially at the corresponding extreme temperatures. Rectangular sections are then a favorable factor (Fig. 24).

4. Materials

Aside from the high-temperature or oxidizing aspect of air-fuel combustion, for a vehicle that has to accelerate and transport a maximum payload, scramjet structures require the lightest possible selection of structural materials. Composite ceramics are preferred.

5. Fuels

As mentioned before, hydrogen is the best fuel in terms of inflammation time, but it is not so in terms of the tank dimensions for this low-density fuel. For the cryogenic engines on today's space launch vehicles, liquid oxygen can be used for civilian vehicles that allow a tank filling phase prior to launch.

Especially for military vehicles, other fuels have to be found. One idea currently being validated is to use a storable liquid fuel having characteristics somewhat like those of kerosene. An endothermic liquid fuel pressurized by a turbine pump, which heats up while circulating through channels in the composite structures of the combustor and cools the combustor at the same time, passes over a catalyst to transform it into gas and is injected in the combustor in gaseous phase with short inflammation delay (Fig. 25).

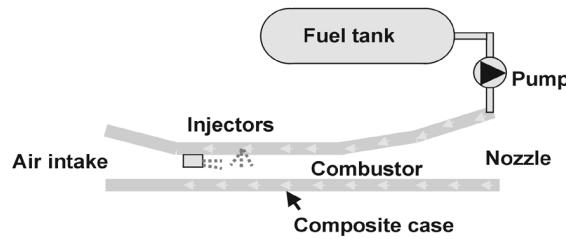


FIGURE 25 Principle of a combustor cooled by an endothermic storable fuel.

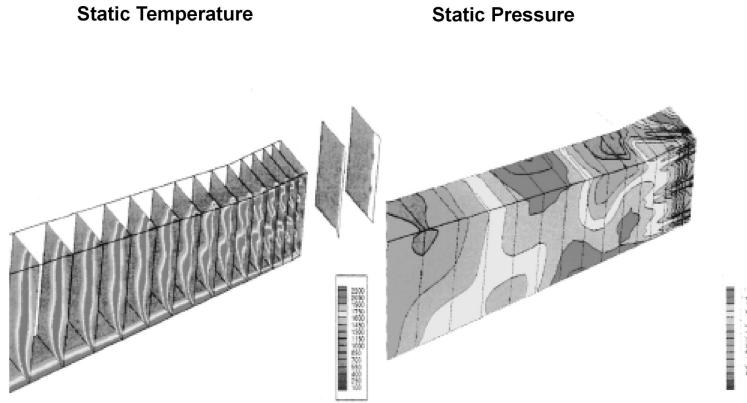


FIGURE 26 Examples of numerical models of combustion in a scramjet.

The combustors currently used in development tests are usually made of copper or steel and are cooled by water circulation. This approach may be usable for short demonstrator flight testing, but never for an operational vehicle.

C. Performances

For reasons of structural strength (pressure and temperature), there is a Mach corridor and flight altitude that hypersonic vehicles with dual mode ramjet propulsion can use. This corridor corresponds to a dynamic pressure of

$$70000 \text{ Pa} < q_0 = \frac{1}{2} \cdot \gamma_0 \cdot p_0 \cdot M_0^2 < 90000 \text{ Pa.}$$

The overall thrust-drag characteristics of scramjet-powered vehicles are not known precisely enough today to undertake a serious project. Ground tests are still needed, as are numerical modeling and flight tests with simplified demonstrators to evaluate the thrust-drag balance more accurately.

Complex numerical modeling is being developed more and more. That is, no simple modeling like the one used for

ramjets is possible, especially considering the presence of numerous shocks and the need to factor in the combustion kinetics.

The best approach for optimizing a combustor is to compare combustion test results with simulation data, which are themselves not always validated.

Figure 26 gives some data generated by numerical combustion models. A number of laboratories have made a great deal of progress over these last few years.

VII. DEVELOPMENT FACILITIES

There are three types of ramjet ground testing facilities:

- The connected pipe setup, without air intakes, feeds directly the combustor with a subsonic flow having a realistic velocity profile measured during wind tunnel tests of the air intakes (Fig. 27).
- The semifree jet setup reproduces the air flow upstream of the air intake when the vehicle is at zero incidence (Fig. 28).
- The free-jet setup reconstructs the real flow around the vehicle and thus also in the air intake area, including

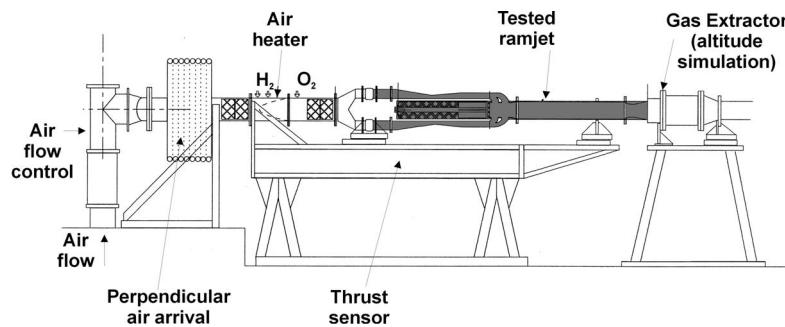


FIGURE 27 Connected pipe test facility for a ramjet.

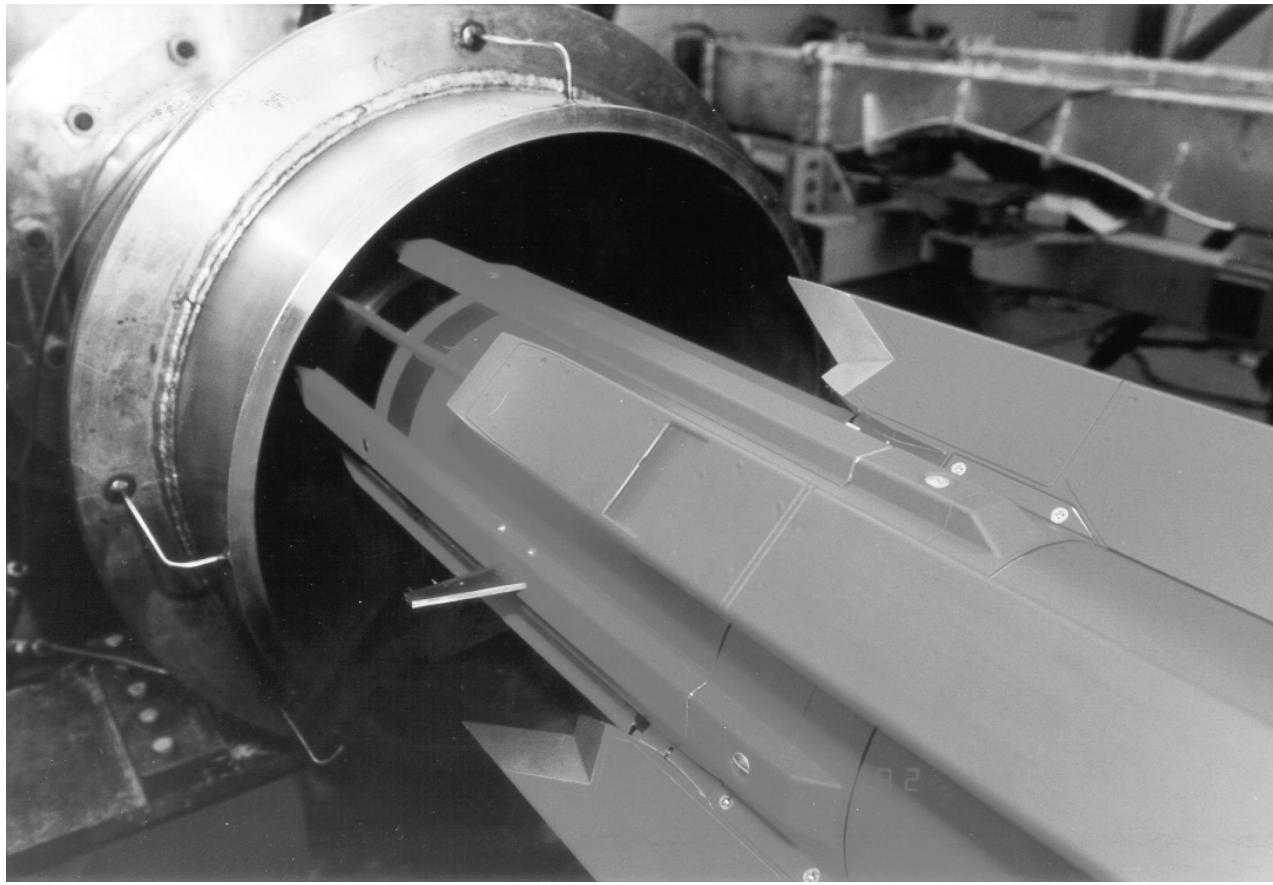


FIGURE 28 View of a semifree jet test facility for a ramjet.

when the vehicle is at an angle of attack, yaw, or even roll.

These last two modes require very costly facilities which, supposing that they exist, are used only after the combustor has been perfected in connected pipe testing.

In all cases, the air has to be supplied at the right mass flow rate and at real stagnation pressures and temperatures. This temperature (ranging from 520 K at Mach 2 at zero altitude to 1700 K or more in hypersonic flight) is generally obtained either by heat exchanger or by air-fuel precombustion. In this last case, oxygen is reintroduced into the flow to reproduce the real air composition. With a hydrogen type fuel there is an excess of water vapor at this point, and with hydrocarbons there is excess water and carbon. For ramjets, it turns out that using hydrogen often has little sensitivity on the combustion efficiency and combustor operating range. For scramjets, though, there is often too much water, and the repercussions of real air on performance then have to be calculated by introducing the true total conditions in a numerical modeling.

The measurement instruments used on the test stands are

- Thermocouples for temperatures
- Pressure transducers, nonsteady if necessary, for detecting any combustion instabilities
- Thrust sensors to measure the exit dynalpy
- Chemical sampling devices
- Optical means for visualizing combustion regions through windows or for determining local chemical element concentrations to compare them with numerical models
- A calorimeter to measure the engine exit energy and further to determine the combustion efficiencies, mainly for scramjets where it is not generally possible to measure the exit dynalpy correctly

Beyond the ground testing, flight measurements are needed to determine the real thrust-drag balance. Though experience shows that ground and flight measurements are in good agreement for ramjet-powered vehicles, there is

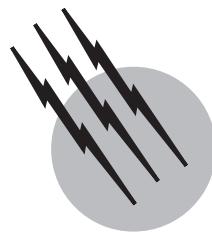
every reason to believe that we are in for some real surprises with scramjets, mainly considering the interaction between the propulsive and external flows, especially at the level of the exhaust.

SEE ALSO THE FOLLOWING ARTICLES

FUELS • JET AND GAS TURBINE ENGINES • LIQUID ROCKET PROPELLANTS • ROCKET MOTORS, HYBRID • ROCKET MOTORS, LIQUID • ROCKET MOTORS, SOLID • SOLID PROPELLANTS • SPACECRAFT CHEMICAL PROPULSION • V/STOL AIRPLANES

BIBLIOGRAPHY

- Cazin, Ph. (1984). "Les statoréacteurs à combustible liquide," ONERA Chatillon, France.
- Jensen, G. E., and Netzer, D. W. (1996). "Tactical Missile Propulsion," AIAA, Reston, Virginia.
- Davenas, A. (1989). "Technologie des propergols solides," Masson, Paris.
- Besser, H., Cochet, A., *et al.* (1994). "Experimental and Analytical Methods for the Determination of Connected-Pipe Ramjet and Ducted Rocket Performances," NATO AGARD/RTO, Neuilly, France.
- Curran, E. T., and Murthy, S. N. B. (2000). "Scramjet Propulsion," AIAA, Reston, Virginia.



V/STOL Airplanes

Barnes W. McCormick

Pennsylvania State University

- I. Introduction
- II. Types of V/STOL Aircraft
- III. Basic Principles of V/STOL Aerodynamics

GLOSSARY

Augmentation Entrainment of a secondary flow by a primary jet to increase thrust for a fixed primary flow.

Boundary layer control Application of a device to prevent flow separation from a surface; usually to increase lift or to decrease drag.

Circulation control Addition of power to provide increased lift on a wing or airfoil.

Disk loading Thrust per unit area of the device producing the thrust.

Downwash Velocity induced downward behind a wing or below a thruster such as a propotor.

Fountain effect Close to the ground, two vertical jets interact to produce an upward flow between them (fountain), which can produce a favorable lift increment on the fuselage.

High-lift device Any device intended to increase the maximum lift that a wing can produce without stalling.

Induced effects Aerodynamic interactions between components of the airplane. For example, the downwash from a wing with a high-lift device can interfere with the tail, greatly reducing the tail's effectiveness.

Induced velocity Velocity caused by the flow reacting to

the production of an aerodynamic force such as lift or thrust.

Lift coefficient Lift of a wing or airfoil expressed in a dimensionless manner (see text for equation).

Lift engine Engine installed in a VTOL airplane for the sole purpose of providing lift.

Lift fan Large, turbine-driven fan mounted in a wing or fuselage to provide vertical lift at low or zero forward speeds.

Payload Useful load carried by an airplane. It includes passengers and cargo but not fuel.

Propotor Rotor that can be tilted downward to serve as a propeller in forward flight.

AIRCRAFT that are capable of taking off or landing in unusually short distances are referred to as short takeoff and landing (STOL) aircraft. An aircraft that can take off or land vertically without the need for any ground roll is known as a vertical takeoff and landing (VTOL) aircraft. Some aircraft are capable of both vertical and short takeoffs and landings. Such aircraft are known as V/STOL aircraft. The acronym STOVL has been coined to refer to a class of airplanes that can take off short but, after dropping their payload or burning off their fuel, are capable of a vertical landing.

I. INTRODUCTION

The takeoff or landing distance for a conventional airplane increases approximately with the square of its takeoff or landing speed. These velocities are nearly equal to the slowest velocity at which the airplane can fly and still remain in the air. At airspeeds less than this velocity, known as the stall velocity, the maximum lift that can be produced by the wing is less than the weight of the airplane. Thus, in order to land or take off in an unusually short distance, an airplane must be capable of generating lift at low airspeeds. Devices to accomplish this are known as high-lift devices. High-lift devices take many forms, from vectoring the exhaust of a turbojet engine downward to blowing air over an elaborate wing-flap system. These are described in more detail in Section II.5.

A VTOL airplane must be capable of generating a lift at zero forward speed equal to, or greater than, its weight. This lift, synonymous with vertical thrust, is produced by accelerating a jet of air downward, producing an equal and opposite force upward. This vertical force can be produced with a small jet of air having a high velocity, such as from a jet engine, or it can be a large jet having a low velocity, such as the flow through a large rotor.

The helicopter is, in a sense, a V/STOL aircraft. However, the performance of a helicopter is limited in forward flight to airspeeds below ~ 200 mph. The helicopter is also inefficient in cruise compared with a fixed-wing aircraft—thus, the desire for the V/STOL airplane having the vertical takeoff and landing capability of a helicopter but with the efficient cruise performance of a conventional takeoff and landing (CTOL) airplane. However, producing a vertical lift with a small jet of fast-moving air requires more power than producing the same lift with a large jet of slow-moving air. Hence, the helicopter is a much more efficient aircraft in hover than the typical V/STOL aircraft. One V/STOL configuration that approaches the hovering capability of a helicopter is the tilt-rotor aircraft to be discussed in Section II.4.

II. TYPES OF V/STOL AIRCRAFT

Many schemes have been studied and flown for V/STOL application. These include the following:

1. *Compound aircraft.* Here, a powered rotor is added to a fixed-wing airplane to achieve VTOL. In forward flight the rotor becomes unloaded as lift develops on the wing. Also, in forward flight, the aerodynamic demands on the rotor are reduced since a separate propulsor, such as a propeller or jet engine, provides the necessary thrust.



FIGURE 1 Tail-sitter XFY-1.

2. *Tail-sitters.* A tail-sitter is, as the name implies, a rather conventional-looking airplane that has enough installed thrust and sufficient controls to hover with its longitudinal axis vertical, that is, with its nose pointed upward. The tail-sitter was the first successful VTOL configuration to fly. The Convair XFY-1 (Fig. 1) was first flown on August 2, 1954. This VTOL airplane was powered by an Allison turboshaft engine developing 5850 shaft horsepower (shp) and driving two coaxial 16-ft-diameter propellers. It had a maximum speed of more than 500 mph. The XFY-1 is now on display at the Smithsonian Institute.

3. *Tilt wing.* This type of V/STOL airplane derives its high lift by rotating the entire wing and propellers through approximately 90° while keeping the fuselage horizontal. The first tilt-wing airplane was the VZ-2 built under U.S. Army sponsorship by the Vertol Co., now a division of the Boeing Co. This flying testbed, powered by a Lycoming YT-53L-1 turboshaft engine, first performed a transition from hover to forward flight in July 1958.

4. *Tilting thruster.* For this class of V/STOL aircraft, the propeller, rotor, or turbojet engine is rotated upward 90° to achieve VTOL performance. It appears, at the time of this writing, that this type of V/STOL, using tilting "proprotors," will evolve as an operational configuration for subsonic V/STOL missions. The term *proprotor* indicates that the large propellers can be rotated 90° upward to serve as rotors for vertical lift. The V-22 Osprey being built jointly by the Boeing Vertol Co. and Bell Helicopter Textron, Inc., is illustrated in Fig. 2. This joint-services advanced vertical-lift program evolved from the XV-15 and the XV-3 airplanes built by Bell Helicopter Textron,

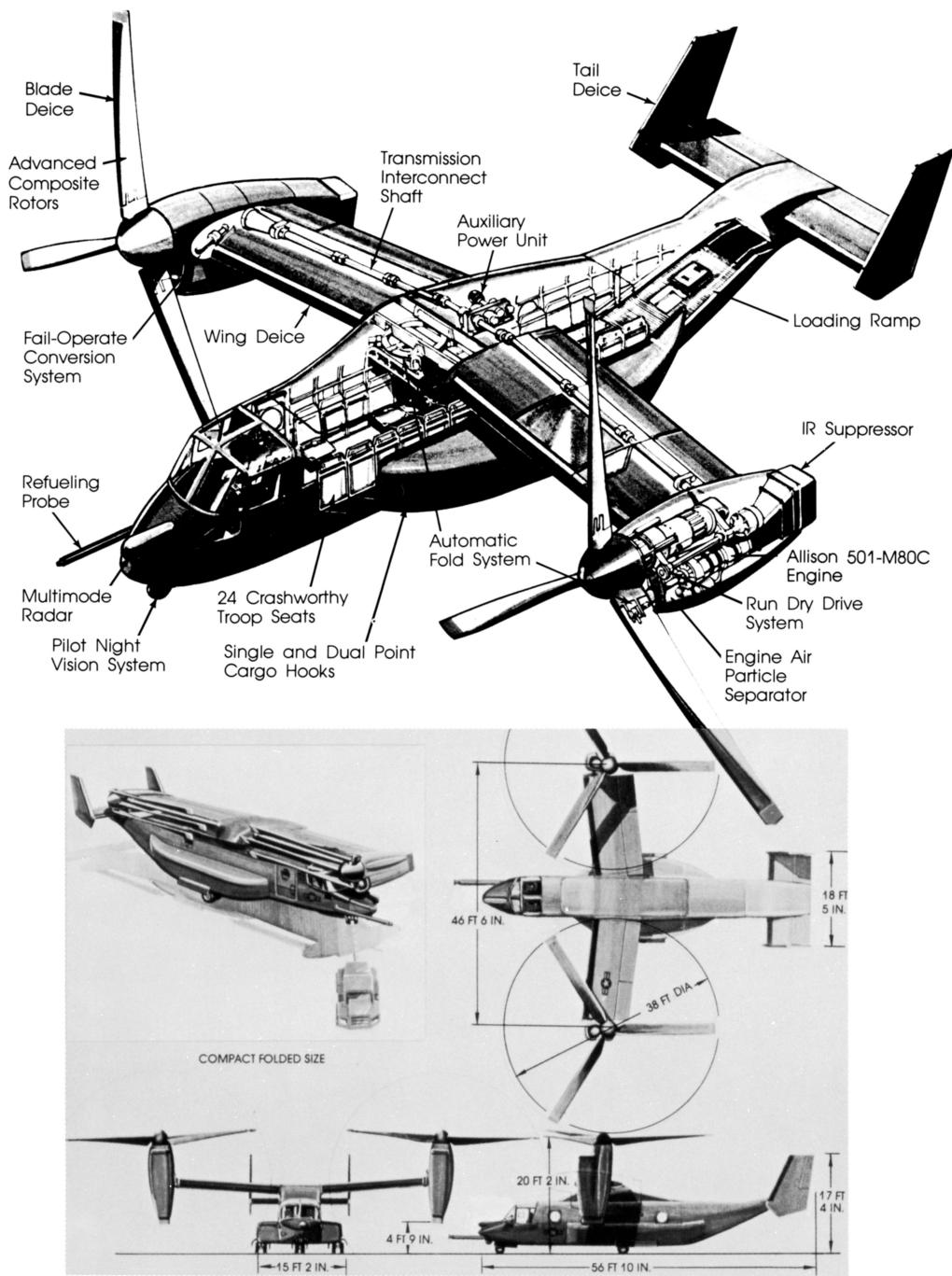


FIGURE 2 Tilting thruster V-22.

Inc. The Bell XV-3, which first flew in 1960, successfully demonstrated the tilt-rotor principle. With the application of modern technology, the XV-15, currently flying, has proved the commercial and military feasibility of the concept. The latest tilt-rotor aircraft is the V-22 Osprey, which flew in 1989. Its performance is illustrated in the graphs of Fig. 3. From the range-payload and time-on-station

graphs it can be seen that the V-22 combines the hovering capability of the helicopter with the forward flight performance of a propeller-driven, fixed-wing airplane.

Another configuration in this class of V/STOL aircraft being actively pursued is the tilt-nacelle, medium-speed V/STOL aircraft designated by the developer, the Grumman Co., as the 698 V/STOL concept. A sketch of

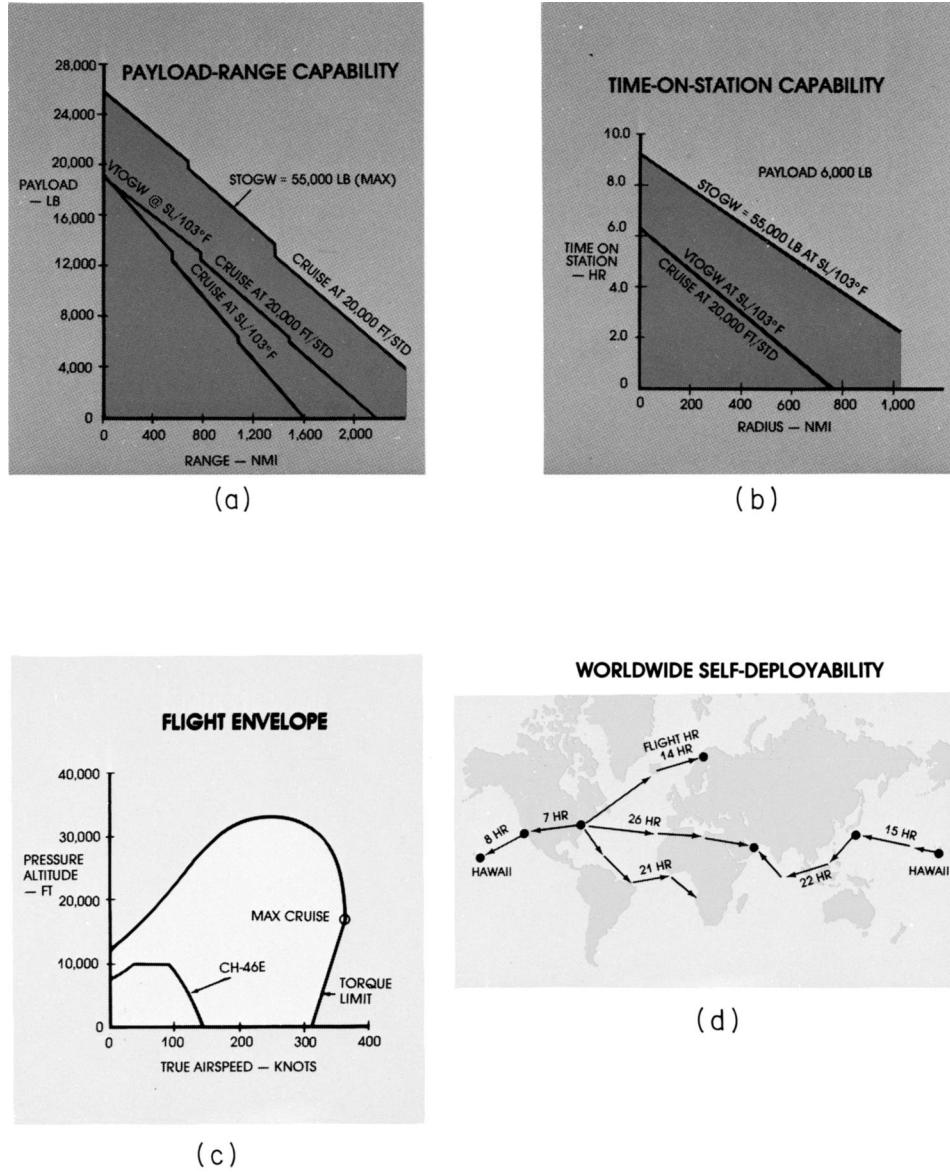


FIGURE 3 Graphs of V-22 performance. (a) Payload-range capability; (b) time-on-station capability; (c) flight envelope; (d) worldwide self-deployability. STOGW: short takeoff gross weight; VTOGW: vertical takeoff gross weight.

this aircraft is shown in Fig. 4. As shown, conventional turbofan engines, with a bypass ratio of approximately 6:1 (typical of modern commercial transports), are tilted to provide vertical thrust. Pitch control of the aircraft is effected by symmetric deflection of horizontal vanes located in the fan exhaust. Similarly, differential deflection of these vanes produces yaw control of the aircraft. For roll control, Grumman has provided vertical vanes in the fan blast as well as variable inlet guide vanes. Approximately 50% of the roll control comes from each source. A major benefit of this system is that the vertical vanes can reduce thrust modulation requirement for the engine for

roll control from approximately ± 20 to $\pm 10\%$ of maximum thrust. This results in significant reduction of the fan diameter and weight of the power plant package.

5. Externally deflected jets. Here, in order to produce lift at low airspeeds, the slipstream from a propeller or the exhaust from a turbojet is deflected downward by a wing-flap system. The earliest of what might be termed a STOL aircraft, the Crouch-Bolas Dragonfly I, flown in 1934, operated on this principle. In recent years, specific configurations in this category have evolved. These are shown schematically in Fig. 5. The externally blown flap (EBF) system was used by the McDonnell Douglas Co. in

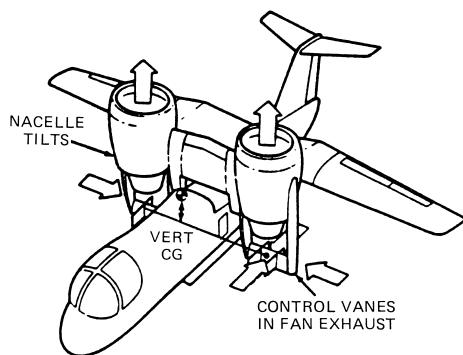


FIGURE 4 Sketch of Grumman 698 V/STOL concept (CG, center of gravity).

the YC-15 STOL transport airplane. The upper-surface-blown (USB) system was employed by the Boeing Co. in the YC-14, another STOL transport. These aircraft, the YC-14 and YC-15, were competing aircraft in a fly-off competition funded by the U.S. Air Force to develop a production STOL airplane. Both airplanes demonstrated excellent maneuverability and short-field performance when flown in the mid-1970s. However, the Air Force terminated the project so that neither design went into production. Subsequently, both Russia and Japan flew transport airplanes using the USB principle. The augmenter wing design has been flown, designated QSRA by the National Aeronautics and Space Administration (NASA) using a modified DeHavilland Buffalo aircraft. In this design, the wing flap deflects to form a nozzle into which the jet engine exhaust is directed. This primary jet entrains a secondary flow, thus increasing the total mass flow of the jet over the flap. QSRA is an acronym for quiet STOL research aircraft.

A major portion of the lift produced by a deflected-jet system derives from the redirection of the momentum in the jet. However, additional lift is produced on the wing by

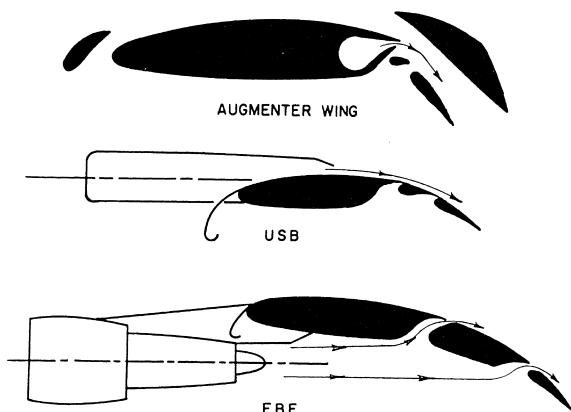


FIGURE 5 Sketches of EBF, USB, and augmenter wing.

the turning of the jet. Thus, there is some overlap between this system and the system to be described next.

6. *Jet flaps and circulation control.* In the case of a jet flap (see Fig. 11), a jet of higher momentum air is ejected from the trailing edge of an airfoil. As the jet trails from the airfoil, the external airstream causes the jet sheet to curve in the direction of the airstream. In order to produce this redirection of the momentum in the jet, a pressure difference must exist across the jet. Hence, unlike an ordinary airfoil, because of the jet flap, a pressure difference can be sustained all the way back on the airfoil to the trailing edge, resulting in an increased lift on the airfoil. This action, whereby the jet controls the pressure distribution on the airfoil, is known as circulation control. However, circulation control is obtained anytime a jet of higher momentum air leaves from the trailing edge of a wing in a downward direction. Thus, the EBF, USB, and augmenter wing configurations discussed previously are closely related to the jet-flap system in their aerodynamic behavior.

Another modification of the jet flap is the blown flap, in which a thin sheet of air exits from a nozzle tangent to the upper surface of a physical flap (see Fig. 11). The jet sheet turns due to the Coanda effect and leaves the airfoil in a direction approximately tangent to the trailing edge of the flap. (The Coanda effect is the action in fluid mechanics whereby a flow along a solid surface tends to follow the curvature of the surface rather than separating—like cream being poured from a pitcher that ends up on the hostess's lace tablecloth instead of in the cup for which it was intended.)

An interesting application of the Coanda effect to obtain circulation control is the elliptically shaped airfoil to be discussed later (see Fig. 12). Air is blown from a slot on the upper rear surface. This jet follows the rounded aft contour around to some point at which the jet separates from the surface. The result is directly comparable to a jet flap. As the jet blowing is increased, the lift increases rapidly because of the increased jet momentum and jet deflection angle. In addition to providing circulation control, the jet prevents the flow generally from separating over the rear of the airfoil, thus achieving a high lift with a low drag.

7. *Internally deflected jet.* The only currently operational VTOL airplane is the highly successful Harrier developed by the Hawker Siddeley Co. of Great Britain. This fighter aircraft uses rotating nozzles, which vector the jet engine exhausts downward to achieve a vertical thrust. A photograph of an improved version of the Harrier, the AV-8B, developed by the McDonnell Douglas Co. is shown in Fig. 6. Also pictured schematically is the rotating nozzle of the Rolls-Royce Pegasus II turbofan engine, which develops 21,550 lb of thrust. The AV-BB has a

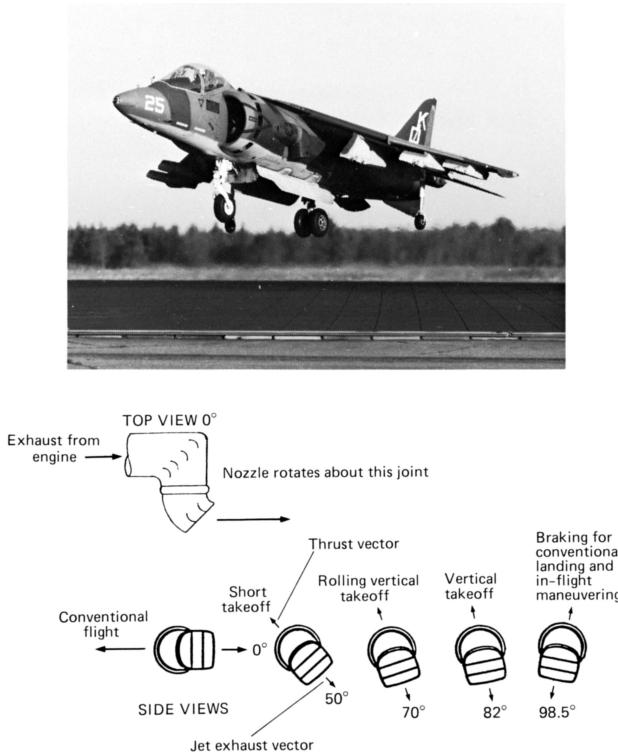


FIGURE 6 Photo of Harrier with sketch of nozzle.

takeoff gross weight of 31,000 lb with an empty weight of 13,086 lb. With a combat radius of 600-plus nautical miles, it can carry 9200 lb of external stores at speeds of up to 562 knots at sea level. For vertical takeoff the nozzle is rotated to point downward. Once airborne, the nozzle is rotated to point to the rear, thus providing the necessary thrust for forward flight.

8. *Direct thrust*. This category includes the use of lift engines and the fan-in-wing. The term *lift engine* means exactly that: a turbojet engine that is mounted on the airplane purely for the purpose of generating vertical thrust. Because of the parasitic weight of such an engine in cruising flight, this concept has not gained acceptance. The fan-in-wing concept was incorporated by the Ryan Airplane Co. in the Ryan XV-5A. This airplane had two fans, one mounted in the left wing and the other in the right wing. A third fan was mounted in the nose, primarily for pitch control. Each fan was powered by a peripheral turbine that was driven by diverting the exhaust of turbojet engines.

9. *Stopped rotor (X-wing)*. The X-wing is, in a sense, a compound aircraft except that for this configuration the rotor is sufficiently strong to be stopped in flight in the manner of an "X," thus providing a forward-swept and aft-swept wing on each side of the aircraft. In order to accomplish this, the rotor employs elliptically shaped airfoils with circulation control (see above) and is thus referred to

as a circulation-controlled rotor (CCR). The use of these unique airfoils allows the rotor blades to be unusually thick without having a high aerodynamic drag. NASA is planning to flight-test the X-wing configuration in the latter half of 1987 to assess the feasibility of the concept. The configuration is rather complex since control is accomplished by cyclically varying the jet momentum issuing from slots located on both the leading and trailing upper surfaces of the blades.

Considerable effort has been devoted to the development of V/STOL aircraft since the 1950s. Most of the configurations that have been studied fall into the above categories. However, to date, only two VTOL configurations have come to fruition. These are the vectored-thrust and tilt-rotor systems represented by the Harrier and the Osprey, respectively. However, currently under development jointly by the United States military services is the Joint Strike Fighter (JSF). The Marine version of this fighter is to have STOVL capability (short takeoff and vertical landing). Two companies, Lockheed Martin and Boeing, are building prototypes for the design competition. The Lockheed Martin version employs a shaft-driven lift fan and a deflected jet for the vertical lift. The Boeing configuration will use rotating nozzles somewhat similar to the Harrier.

III. BASIC PRINCIPLES OF V/STOL AERODYNAMICS

A. Static Thrust

In the design of a VTOL aircraft the division between hovering and cruising flight in the application of the aircraft plays a dominant role in determining the aircraft's configuration. This statement is easily understood from basic momentum principles of fluid mechanics. A static lift system, such as a hovering VTOL aircraft, having a mass flow rate of m through the system and producing a thrust of T will, on the average, accelerate the flow to a final velocity of w . With m , T , and w in consistent units, these quantities are related simply by

$$T = mw$$

The power P delivered to the mass flow by the lift system will equal the flux of kinetic energy in the ultimate wake downstream of the thruster. Thus,

$$P = mw^2/2$$

Thus, combining the above, the power, thrust, and average induced velocity in the ultimate wake are related by

$$T/P = 2/w$$

The above is actually an ideal upper limit on the thrust-to-power ratio and can never be attained in practice because of other power losses in the system. Nevertheless, this equation provides a relative basis for comparing static thrust systems and suggests, for the same thrust, that the fuel flow rate will vary directly with the induced velocity. Typically, w is equal approximately to 40 fps for helicopters, 230 fps for propellers, 590 fps for lift fans, and 980 fps for turbojets. Thus, direct lift systems, on the one extreme, will burn fuel at a rate approximately 25 times that of the helicopter at the other end of the w spectrum. On the other hand, the weight of a direct lift turbojet system will generally be less than that of a system with a low w (e.g., a helicopter with its large blades, hub, transmissions, etc.) so that weight can be traded for fuel.

For a given thrust, an important parameter that governs the induced velocity is the disk loading. The disk loading is defined as the ratio of the thrust to the projected area of the thruster over which the thrust is being produced. Thus, the disk loading represents the average pressure increase across the area swept by the thruster. For a propeller having a radius R and producing a thrust T , the disk loading is T/A , where A is the disk area, πR^2 . It is easily shown that the velocity induced through a propeller is equal to $w/2$. Thus, the mass flow m through the propeller can be calculated from

$$m = \rho A w / 2$$

where ρ is the mass density of the air. Substituting the above for m into the equation for the thrust results in a relationship between w and the disk loading, namely,

$$w = (2T/\rho A)^{1/2}$$

These relationships, presented graphically in Fig. 7, show the strong dependence of the power loading T/P on the disk loading. Typical ranges for various types of VTOL aircraft are indicated in the figure.

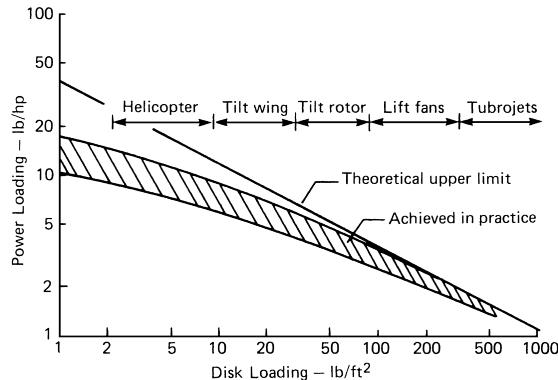


FIGURE 7 Effect of disk loading on power loading.

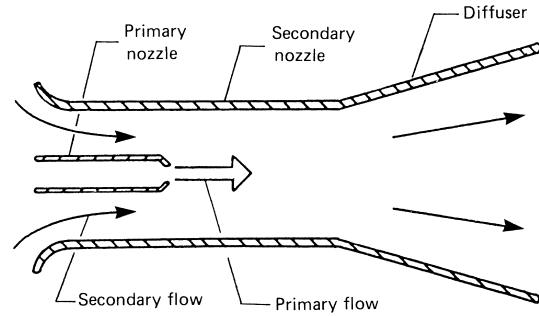


FIGURE 8 Sketch of augmenter.

1. Augmentation

One means of reducing the induced velocity in the ultimate wake is to use augmentation. As shown schematically in Fig. 8, a primary jet is introduced into the throat of a surrounding nozzle, either along the centerline of the nozzle or along its walls. Viscous shear and turbulent mixing along the boundary of the primary jet entrain a secondary flow through the surrounding nozzle. This secondary flow mixes with the primary jet to produce an augmented mass and momentum flux issuing from the nozzle, which can be appreciably greater than that of the primary nozzle. The performance of the system can be improved by a diffusing shroud downstream of the secondary nozzle and also by the application of devices (e.g., hypermixing nozzles) to provide better mixing of the primary and secondary flows. Unfortunately, the results from flight testing of augmenters have not been as good as the performance predicted from analysis and laboratory testing. The few VTOL airplanes in which augmentation was applied to increase performance have not been very successful.

2. Fountain Effect and Suckdown

A sketch of two jets issuing from beneath the fuselage of a VTOL airplane is shown in Fig. 9. Close to the ground these jets interact with the ground and tend to spread outward. However, as the flow from one jet approaches the flow from the other, an upward jet is formed resembling a fountain, hence the term *fountain effect*. As the fountain between the two jets impinges on the bottom of the fuselage, it produces an increased pressure in this region, which causes added lift; hence, the fountain effect is a favorable one.

Looking end-on at the two-jet system of Fig. 9, it is seen that the two jets entrain a secondary flow vertically downward around the fuselage. As this flow curves around the bottom edges of the fuselage, the pressure is reduced there. This lower pressure acting over the bottom of the fuselage detracts from the lift; thus, this effect, called *suckdown*, is an adverse one. It can be alleviated by mounting vertical

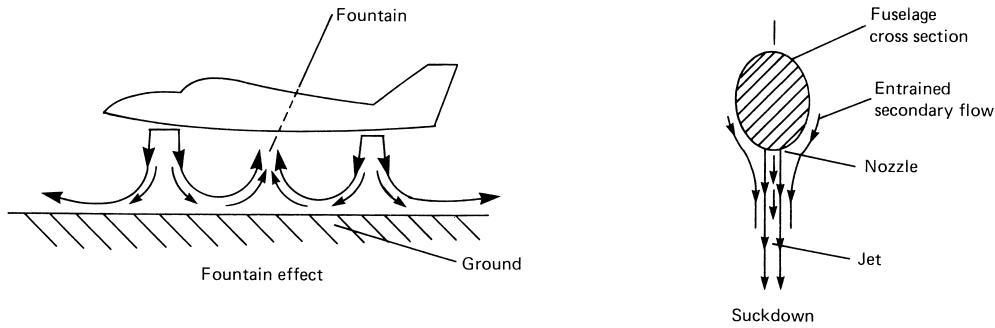


FIGURE 9 Sketch showing fountain effect and suckdown.

plates along the sides and extending below the bottom of the fuselage to prevent the entrained flow from curving around to the bottom. Such plates are called strakes.

The net result of the fountain effect and suckdown can be either favorable or undesirable, depending on the height of the fuselage above the ground and the fuselage geometry. Test results for the Grumman 698 V/STOL concept are presented in Fig. 10. This figure clearly shows the influence of the ground and strakes on the net induced force as a fraction of the gross thrust. With no strakes the thrust is reduced due to suckdown at heights of up to four jet diameters above the ground. With strakes deflected 45° outward the fountain effect results in an 11% increase in the thrust at a height of one jet diameter above the ground.

3. Glauert's Hypothesis

Glauert's hypothesis states that, approximately, the induced effects for a lifting system in forward flight (or statically) can be determined by applying momentum principles to an area enclosed by a circle that circumscribes

the system. This area is taken normally to the resultant velocity through the system. This hypothesis agrees exactly with the results for a statically thrusting propeller presented earlier where the thrust vector is aligned with the flow through the propeller. It can also be shown to agree with a lifting wing where the thrust vector (the lift) is perpendicular to the velocity past the lifting wing. Since it holds at these two limits, it is hypothesized that this approximate application of momentum principles will hold for cases in between where the thrust vector is at some arbitrary angle to the resultant velocity through the system.

Glauert's hypothesis was the basis for analyzing much of the aerodynamics of the XV-15 and V-22 (shown in Fig. 2). A rotor is shown at a high angle of attack in Fig. 11. V denotes the forward velocity of the tilt-rotor airplane and w the velocity induced by the thrust. This velocity is opposite in direction to the thrust. Using these velocities, the resultant velocity through the rotor is obtained as;

$$V_R = \sqrt{(V + w \cos \alpha)^2 + (w \sin \alpha)^2}$$

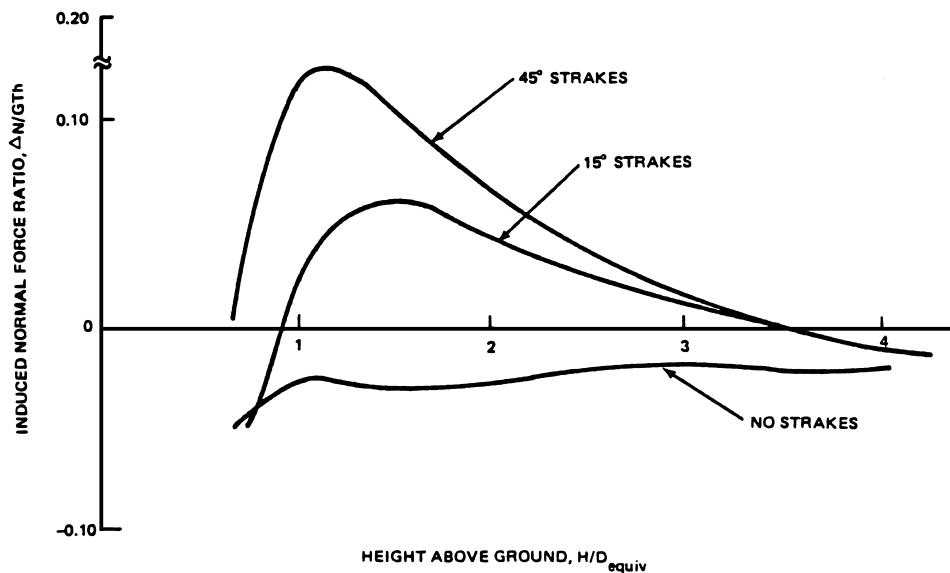


FIGURE 10 Suckdown and fountain effect for Grumman 698.

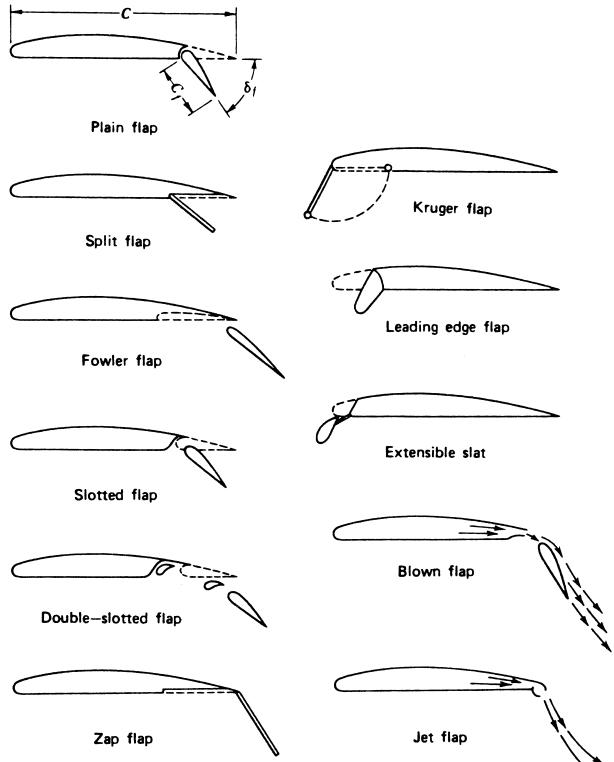


FIGURE 11 Drawings of flaps.

The thrust is then related to V_R by;

$$T = 2\rho AV_R w$$

The ideal power required to produce the thrust inclined at some arbitrary, α , is given by,

$$P = T(V \cos \alpha + w)$$

At low forward speed, the above equation for the power represents most of the power. To this must be added an increment to account for the power required to overcome the drag of the rotor blades. The details for solving the equations resulting from the application of Glauert's hypothesis can be found in the reference by McCormick. This hypothesis can also be applied to helicopters, particularly when climbing or descending where the angle of attack of the rotor is high.

B. Aerodynamics of High Lift

The lift of an airplane wing is proportional to the product of the air mass density ρ , the wing area S , and the square of the airspeed V . The constant of proportionality is called the wing lift coefficient and is denoted by C_L . This relationship for the lift L can be written

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

Here, the quantity $\frac{1}{2} \rho V^2$ is referred to as the dynamic pressure.

The wing lift coefficient C_L is dependent on the wing geometry and its angle relative to the airflow. For a given wing shape, C_L increases linearly with this angle, denoted as the angle of attack, up to a maximum value $C_{L_{\max}}$, beyond which the flow separates from the upper surface of the wing, causing a loss of lift. Here, the wing is said to be stalled. In steady, level flight, the lift of the wing must equal the weight W of the airplane. Thus, it follows that the slowest speed at which an airplane can fly and be supported by the wing lift, the stalling speed V_s , is given by $V_s = (2W/\rho S C_{L_{\max}})^{1/2}$. To achieve good STOL performance, an airplane must therefore produce a high $C_{L_{\max}}$. Typically, an ordinary airfoil will produce a maximum lift coefficient of ~ 1.6 . To increase $C_{L_{\max}}$, movable surfaces known as flaps are fitted to the leading and trailing edges of the airfoil. Various configurations of leading and trailing edge flaps are shown in Fig. 11. Some of the flaps increase the chord of the airfoil when they are extended. An elaborate flap system, such as a double-slotted trailing edge flap combined with a Krueger leading edge flap, can achieve $C_{L_{\max}}$ values of ~ 3.0 .

In order to achieve $C_{L_{\max}}$ values higher than those possible with unpowered flap systems, one must utilize power. Here, the airflow around the airfoil is energized so as to delay flow separation or increase the pressure difference (or both), and hence the lift, from the upper surface of the airfoil to the lower. The product of the airfoil chord and the mean velocity over the upper surface minus that over the lower surface is referred to by the aerodynamicist as the circulation Γ . Stated precisely, the circulation is defined as the closed line integral about the airfoil of the velocity vector tangent to the direction of integration. A well-known theorem, the Kutta-Joukowski theorem, states that the lift per unit span on an airfoil is related to the circulation by

$$L = \rho V \Gamma$$

The jet flap and other similar high-lift devices increase the circulation around an airfoil to values higher than can be achieved without power. This action caused by the trailing sheet of high momentum air was defined earlier as circulation control.

A parameter that characterizes the performance of a jet flap is the momentum coefficient C_μ . This coefficient is equal to the momentum in the jet per unit wing area divided by the free-stream dynamic pressure:

$$C_\mu = m_j v_j / q S$$

Here, m_j is the mass flow rate from the jet having a velocity of v_j , S is the planform area of the wing, and q represents the dynamic pressure.

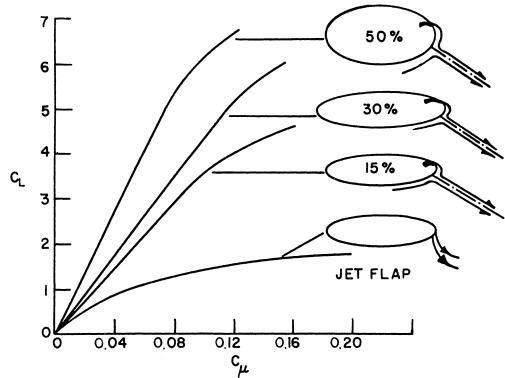


FIGURE 12 Graph of lift performance of elliptical airfoils with circulation control.

The lift performance of the elliptically shaped airfoils discussed previously and being used on the X-wing CCR rotor is presented in Fig. 12. The bottom curve shows the lift coefficient as a function of C_μ for a fixed jet-flap deflection angle. Here, the behavior is similar to an ordinary jet flap. For the other curves, the jet is free to separate at the point determined by the jet momentum. As a result, the deflection angle of the jet increases with C_μ , giving a faster rate of increase of C_L with C_μ as compared with

the jet-flap case. Observe that C_L values more than double those attainable without power are achieved with relatively low values of C . Of course, this increase in the lift coefficient requires some expenditure of power as well as some additional weight for the system to duct the air from the compressor. The power can be determined from the flux of kinetic energy in the jet and a knowledge of the compressor efficiency and losses in the air ducts.

SEE ALSO THE FOLLOWING ARTICLES

AIRCRAFT SPEED AND ALTITUDE • AIRCRAFT PERFORMANCE AND DESIGN • AIRPLANES, LIGHT • FLIGHT (AERODYNAMICS) • HELICOPTERS

BIBLIOGRAPHY

- Campbell, J. (1962). "Vertical Takeoff and Landing Aircraft," Macmillan, New York.
- Kohlman, D. L. (1981). "Introduction to V/STOL Airplanes," Iowa State Univ. Press, Ames.
- McCormick, B. W. (1967). "Aerodynamics of V/STOL Flight," Academic Press, New York.