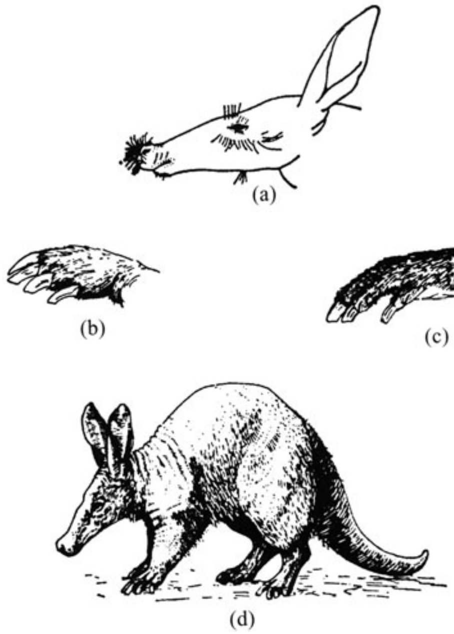
A

AAAS. The American Association for the Advancement of Science was founded in 1848 and incorporated in 1874. Its objectives are to further the work of scientists, to facilitate cooperation among them, to foster scientific freedom and responsibility, to improve the effectiveness of science in promoting human welfare, to advance education in science, and to increase public understanding and appreciation for the importance and promise of the methods of science in human progress. The AAAS headquarters is in Washington, D.C.

AARDVARK (*Mammalia, Tubulidentata*). African animals of peculiar form and ancient lineage, including an Ethiopian and a South African species. All are anteaters, feeding exclusively on ants and termites, nocturnal in habit, with acute hearing. The southern species has been called the ant bear. The aardvark is the only living representative of its order. The animal's spine, curved from neck to tail in a near-half circle, gives it a truly prehistoric appearance.

The aardvark is solitary. In daytime it sleeps curled up like a dog in one of its burrows, often beneath a termite hill. The animal moves almost entirely at night, when it seeks termite hills and destroys them to reach the interior chambers and tunnels alive with insects, which it rapidly licks up in lumps. Although usually silent, the animal can grunt like a hippopotamus. Life span is at least 10 years. In captivity, aardvarks become accustomed to keepers, but do not show great intelligence. See accompanying illustration.

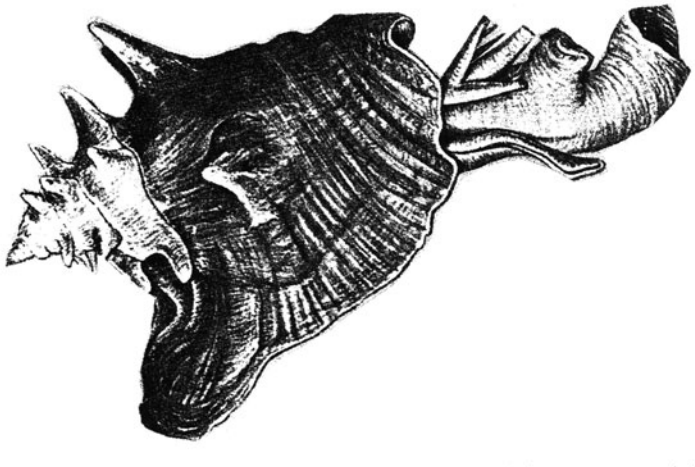


Aardvark, a sturdy animal about the size of a pig. (a) A long conical head bears small eyes and large ears, each of the latter moving independently. The short neck leads to a round high-backed rump with a long stout tail and short stout legs. The snout terminates with a kind of trunk-shaped snout, like a pig's. The mouth is narrow and allows only a long, strap-like, adhesive tongue to pass. The color ranges between yellow-gray and brown, especially the legs show often a darker brown or even black color. The forefeet (b) show four fingers only; the thumb or first finger has vanished completely. The hind feet (c) have five toes each, with broad claws and webbing between the second and the fourth toes. The animal is able to erect the forebody or to sit down in erected position. (d) The habitus or body definitely purveys a pig-like appearance. (*Sketches after Grasse, 1955.*)

ABACA. The sclerenchyma bundles from the sheathing leaf bases of *Musa textilis*, a plant closely resembling the edible banana plant. These bundles are stripped by hand, after which they are cleaned by drawing over a rough knife. The fiber bundles are now whitish and lustrous, and from six to twelve feet (1.8–3.6 meters) long. Being coarse, extremely strong and capable of resisting tension, they are much used in the manufacture of ropes and cables. Since the fibers swell only slightly when wet, they are particularly suited for rope which will be used in water. Waste manila fibers from rope manufacture and other sources are used in the making of a very tough grade of paper, known as manila paper. The fibers may be obtained from both wild and cultivated plants, the latter yielding a product of better grade. The cultivated plants, propagated by seeds, by cuttings of the thick *rhizomes* or by suckers, are ready for harvest at the end of three years, after which a crop may be expected approximately every three years.

ABALONE (*Mollusca, Gasteropoda; Haliotis*). Marine species, mostly of the Pacific and Indian Oceans. The single broad shallow shell has a richly colored iridescent inner surface and is an important source of mother-of-pearl and blister pearls for costume jewelry. The flesh is palatable.

The few whorls of abalone shells are flattened and rapidly increase in diameter so that the largest part of the shell consists of the last part of the whorl (*ear-form*). The European representative of this family is the *Ormer* (*Haliotis tuberculata*), which is characterized by knotty longitudinal shell ridges. It is distributed from the English Channel to western Africa. *Haliotis tuberculata lamellosa* (accompanying figure) is characterized by transverse shell ridges. Both forms usually measure 5 to 7 centimeters. The largest species occur along the Pacific Coast of the United States, in northern Japan, and in southern Australia, where the animals are commercially fished for their delicious meat. The iridescent shells are used extensively in making jewelry. Most of the approximately seventy species are found in cool waters. Some members of this group attain shell lengths of over 20 centimeters and live to an



Abalone (*Haliotis tuberculata lamelosa*), characterized by transverse shell ridges.

D. M. Considine et al. (eds.), *Van Nostrand's Scientific Encyclopedia*

age of 10 to 13 years. Aside from the characteristics already mentioned, abalones are further unique in that the slit band is present only as a series of small holes. During further growth, abalones develop additional holes in a curved line, and the "retired" holes are sealed over.

The abalones inhabiting the intertidal zone to a depth of about 50 meters scrape algae off rocks. All abalone species have adapted to this ecological niche by developing a broad suction foot with a correspondingly large shell. The light-shunning animals attach themselves to shady parts of the rock with this broad foot. The suction force of this foot is more than 4000 times that of the animal's body weight. See also **Mollusks.**

ABBE CONDENSER. A compound lens used for directing light through the object of a compound microscope. All the light enters the object at an angle with the axis of the microscrope. See also **Microscope.**

ABBE NUMBER. The reciprocal of the dispersive power of a material. It is also called the ν-number.

ABBE SINE CONDITION. The relationship

$$ny \sin \theta = n'y' \sin \theta',$$

where n,n' are refractive indices, y,y' are distances from optical axis, and θ,θ' are angles light rays make with the optical axis. A failure of an optical surface to satisfy the sine condition is a measure of the coma of the surface.

ABDOMEN. The abdomen is the posterior division of the body in many arthropods. It is the *posterior* portion of the trunk in vertebrates. In the vertebrates this region of the body contains most of the alimentary tract, the excretory system, and the reproductive organs. It contains part of the coelom and in mammals is separated from the thorax by the diaphragm.

The abdominal cavity of the human body is subdivided into the abdomen proper and the pelvic cavity.

The walls of the abdominal cavity are lined with a smooth membrane called the peritoneum, which also provides partial or complete covering for the organs within the cavity.

The abdomen proper is bounded above by the diaphragm; below it is continuous with the pelvic cavity; posteriorly it is bounded by the spinal column, and the back muscles; and on each side by muscles and the lower portion of the ribs. In front, the abdominal wall is made up of layers of fascia and muscles. The abdomen is divided into nine regions whose boundaries may be indicated by lines drawn on the surface. The mid-section above the navel between the angle of the ribs is known as the epigastric region; that portion around the navel, as the umbilical; below the navel and above the pubic bone, as the hypogastric region. It is further divided into right and left upper quadrants on each side above the navel, and right and left lower quadrants on each side below the navel. The lumbar region extends on either side of the navel posteriorly and laterally.

The principal organs of the abdominal cavity are the stomach, duodenum, jejunum, ileum, and colon or large intestine, the liver, gall bladder and biliary system, the spleen, pancreas and their blood and lymphatic vessels, lymph glands, and nerves, the kidneys and ureters.

The pelvic portion of the abdomen contains the sigmoid colon and rectum, a portion of the small intestine, the bladder, in the male the prostate gland and seminal vesicles, in the female the uterus, Fallopian tubes and ovaries.

ABEL EQUATION. A mass point moves along a smooth curve in a vetical plane and under the influence of gravity alone. Given the time, t, required for the particle to fall from a point, x, to the lowest point on the curve as a function of x, what is the equation of the curve? The problem leads to a Volterra integral equation of the first kind.

$$f(x) = \int_0^x \frac{\phi(t)\, dt}{\sqrt{2g\,(x-t)}}$$

where g is the acceleration of gravity. The solution is

$$\phi(x) = \frac{\sqrt{2g}}{\pi} \int_0^x \frac{f'(t)\, dt}{\sqrt{x-t}}$$

and the equation of the curve is

$$y = \int_0^x \sqrt{|\phi^2(t) - 1|}\, dt$$

A closely related problem is that of the brachistochrone, where the path is required for a minimum time of descent. Such matters were of considerable interest to many seventeenth and eighteenth century mathematicians; the one described here was solved by the Norwegian, N. H. Abel (1802–1829). See also **Brachistochrone.**

A more general case of the Abel equation is

$$f(x) = \int_0^x (x-y)^{-\alpha} \phi(y)\, dy$$

where $f(x)$ is continuously differentiable for $x \geq 0$ and $0 < \alpha < 1$. The solution is

$$\phi(y) = \frac{\sin \alpha\pi}{\pi} \left[\int_0^y (y-x)^{\alpha-1} f'(x)\, dx + f(0) y^{\alpha-1} \right]$$

A first-order differential equation

$$y' = f_0(x) + f_1(x)y + f_2(x)y^2 + f_3(x)y^3$$

is also known as an Abel equation. When the $f_i(x)$ are given explicitly, the equation can often be converted into one of simpler type and solved in terms of elementary functions. In the general case the solution involves elliptic functions.

ABELIAN GROUP. A commutative group, namely such that $AB = BA$ where A, B are any two elements contained in it.

ABERRATION OF LIGHT. The apparent change of position of an object, due to the speed of motion of the observer. Care must be taken not to confuse this effect with that of parallax.

If a telescope, assumed to be stationary, is pointed at a source of light, the light that enters the object glass centrally and in the direction of the optic axis will pass through the telescope along that axis and emerge through the center of the eyepiece. If the telescope is in motion relative to the source, in any direction other than parallel to the optic axis, the light that enters centrally will emerge off the center of the eyepiece. If this light is to emerge centrally, the telescope must be tilted forward in the plane containing the direction of motion of the instrument and the source. The amount of tilt will depend on the direction of the source and the ratio of the speed of the telescope to the speed of light.

This aberrational effect was first announced by Bradley in 1726. He noticed that stars had apparent periodic motions with a period of one sidereal year, and that the character of the apparent motion depended upon the celestial latitude of the star. He correctly interpreted the effect as due to the motion of the earth about the sun. Statistical discussions of the observations of a large number of stars have shown that the maximum value of this aberration due to the earth's orbital motion is $20''.47$. This is known as the "aberration angle" or the "constant of aberration," and is given by

$$\kappa = \frac{2\pi a \operatorname{cosec} 1''}{cT(1-e^2)^{1/2}}$$

where a is the mean radius of the earth's orbit, c is the velocity of light, T is the length of the year in seconds, and e is the eccentricity of the orbit. An aberrational effect of about $0''.3$, at maximum, is observed, due to the rotation of the earth on its axis, and is given by

$$k = \frac{2\pi\rho \cos \phi \operatorname{cosec} 1''}{ct}$$

where ρ is the radius of the earth, ϕ is the latitude of the place, and t is the length of the day in seconds.

In 1871, Airy made a series of observations for determination of the aberration constant, using a telescope filled with water. Because the value of the index of refraction of water is about $1\frac{1}{3}$, Airy expected that the value of the aberration would be 27″.3 when using the water-filled tube. He found, however, that the value was 20″.5 no matter what substance was placed in the telescope. The result of this so-called "Airy's Experiment" caused much discussion, but was eventually explained on the basis of the Michelson-Morley experiment and the theory of relativity.

All observations in which the positions of the stars are involved must be corrected for aberration of light if the results are to be accurate to within 20″. Both the motion of the earth about the sun and the rotation of the earth must be considered. The magnitude of the correction depends upon the celestial coordinates of the star, the position of the observer on the earth, and the date and time of observation.

ABERRATION (Optical). The failure of an optical system to form an image of a point as a point, of a straight line as a straight line, and of an angle as an equal angle. See also **Astigmatism; Chromatic Aberration; Coma (Optics); Curvature of Field (Optics); Spherical Aberration.**

ABLATION (Geomorphology). Essentially, the wasting away of rocks; the separation of rock material and formation of residual deposits, as caused by wind action or the washing away of loose and soluble materials.

ABLATION (Glaciology). The combined processes (sublimation, melting, evaporation) by which snow or ice is removed from the surface of a glacier or snowfield. In this sense, the opposite of alimentation. Ablation also refers to the amount of snow or ice removed by the aforementioned processes (the opposite of accumulation). The term may be applied to reduction of the entire snow-ice mass, and may also include losses by wind action and by calving (the breaking off of ice masses). Air temperature is the dominant factor in controlling ablation. During the ablation season, an ablation rate of about two millimeters/hour is typical of most glaciers. An ablatograph is an instrument that measures the distance through which the surface of snow, ice, or firn charges, as caused by ablation, during a specific period.

ABLATION (Meteorite). The direct vaporization of molten surface layers of meteorites and tektites during flight.

ABLATION (Spacecraft). In the interest of cooling space vehicles upon re-entry into the earth's atmosphere, ablation is used to control the temperature of strongly heated surfaces, such as parts of combustion chambers or nose cones. The process usually consists of the use of surface layers of materials which by their fusion, followed often by evaporation, absorb heat.

The heat of ablation is a measure of the effective heat capacity of an ablating material. Numerically, this is the heating rate input divided by the mass loss rate which results from ablation. In the most general case, heat of ablation is given by

$$(q_c + q_r - \sigma\epsilon T_w^4)/\mathring{m}$$

where q_c is the convective heat transfer in the absence of ablation; q_r is the radiative heat transfer from hot gases to ablation material; $\sigma\epsilon T_w^4$ is the rate of heat rejection by radiation from external surface of ablation material; and $\mathring{m}$ is rate at which gaseous ablation products are injected into the boundary layer.

Heat of ablation is sometimes evaluated neglecting the heat rejected by radiation and as a result unrealistically high heats of ablation are obtained.

If $q_r < \sigma\epsilon T_w^4$, for moderate values of stream enthalpy h_s, the heat of ablation is given by

$$H_v + \eta(h_s - h_w)$$

where H_v is the heat required to cause a unit weight of mass to be injected into boundary layer; η is the blocking factor, with numerical value from about 0.2 to 0.6 depending on material and type of flow; and h_w is the enthalpy at wall temperature.

ABNEY EFFECT. A shift in hue which is the result of a variation in purity and, therefore, in saturation. The Abney effect may be represented by chromaticity loci, of specified luminance, with the hue and brightness constant, when purity and, therefore, saturation are varied. It is a relationship, of psychophysical nature, between psychophysical specifications and color sensation attributes.

ABRASION. All metallic and nonmetallic surfaces, no matter how smooth, consist of minute serrations and ridges which induce a cutting or tearing action when two surfaces in contact move with respect to each other. This wearing of the surfaces is termed abrasion. Undesirable abrasion may occur in bearings and other machine elements, but abrasion is also adapted to surface finishing and machining, where the material is too hard to be cut by other means, or where precision is a primary requisite.

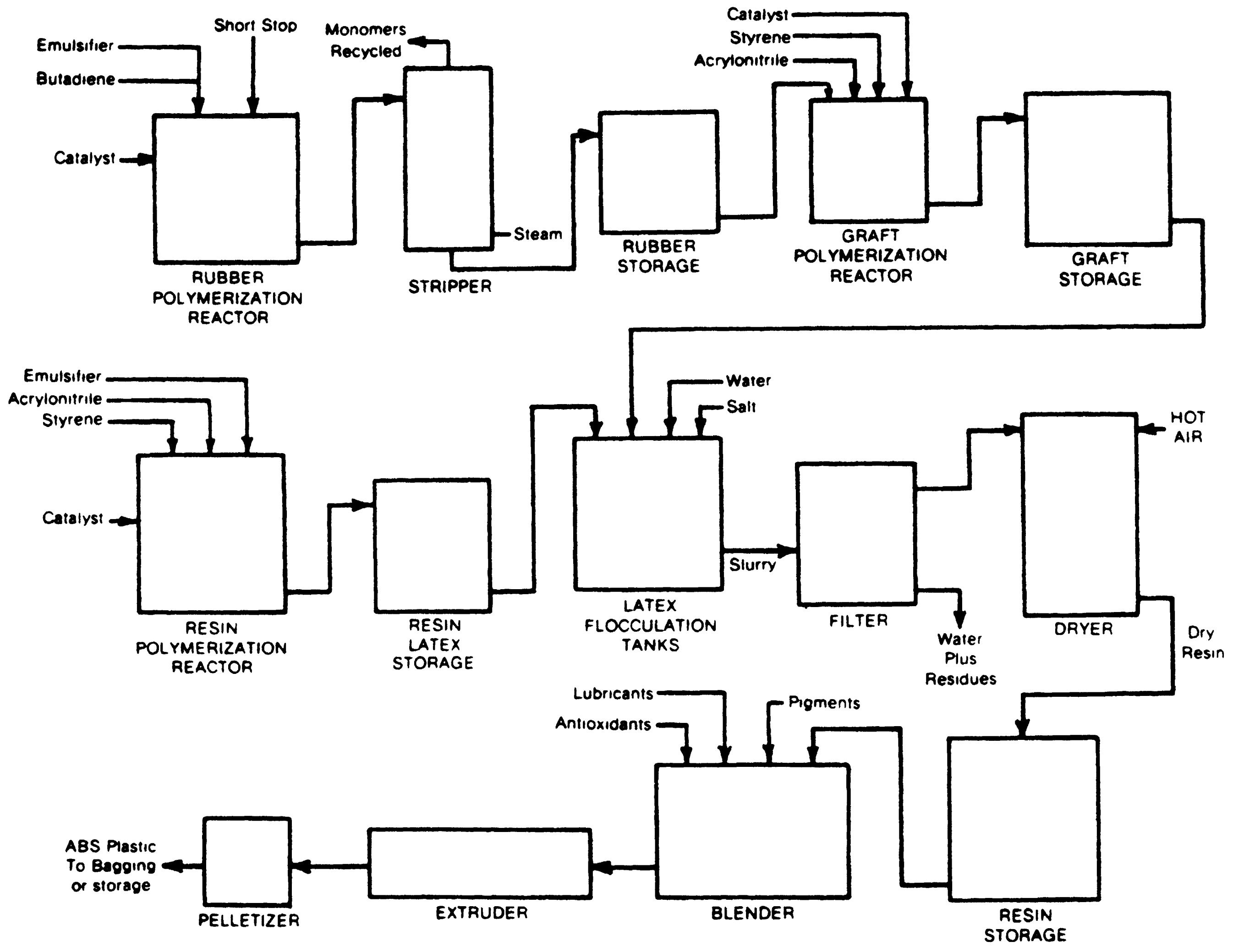
ABRASION pH. A term originated by Stevens and Carron in 1948 "to designate the pH values obtained by grinding minerals in water." Abrasion pH measurements are useful in the field identification of minerals. The pH values range from 1 for ferric sulfate minerals, such as coquimbite, konelite, and rhomboclase, to 12 for calcium-sodium carbonates, such as gaylussite, pirssonite, and shortite. The recommended technique for determining abrasion pH is to grind, in a nonreactive mortar, a small amount of the mineral in a few drops of water for about one minute. Usually, a pH test paper is used. Values obtained in this manner are given in the left-hand column of the accompanying table. Another method, proposed by Keller et al. in 1963 involves the grinding of 10 grams of crushed mineral in 100 milliliters of water and noting the pH of the resulting slurry electronically. Values obtained in this manner are given in the right-hand column of the accompanying table.

ABRASION pH VALUES OF REPRESENTATIVE MINERALS

Mineral	pH by Stevens-Carron Method	pH by Keller et al. Method*
Coquimbite	1	
Melanterite	2	
Alum	3	
Glauconite	5	5.5*
Kaolinite	5, 6, 7	5.5*
Anhydrite	6	
Barite	6	
Gypsum	6	
Quartz	6, 7	6.5
Muscovite	7, 8	8.0
Calcite	8	8.4
Biotite	8, 9	8.5
Microcline	8, 9	8.0 9.0*
Labradorite		8.0 9.2*
Albite	9, 10	
Dolomite	9, 10	8.5
Hornblende	10	8.9
Leucite	10	
Diopside	10, 11	9.9
Olivine	10, 11	9.6*
Magnesite	10, 11	

*indicates more recent values published in literature.

ABS (Acrylonitrile-Butadiene-Styrene) RESINS. Thermoplastic resins which are produced by grafting styrene and acrylonitrile onto a diene-rubber backbone. The usually preferred substrate is polybutadiene because of its low glass-transition temperature (just above −80°C). Where ABS resin is prepared by suspension or mass polymerization methods, stereospecific diene rubber made by solution polymerization



Acrylonitrile-butadiene-styrene (ABS) resin manufacturing process. (*BorgWarner Chemicals.*)

is the preferred diene. Otherwise the diene used normally is a high-gel or cross-linked latex made by a "hot-emulsion" process.

In most ABS polymerization approaches, a rubber substrate is manufactured in an initial stage, with the styrene/acrylonitrile copolymer matrix being generated in one or more subsequent stages. An example of one approach is shown in block diagram format in the accompanying illustration.

STEP 1—Polybutadiene rubber is formulated by feeding butadiene, water, an emulsifier, and catalyst into a glass-lined reactor. This is an exothermic reaction. About 80% conversion is achieved in a period of about 50 hours. The residual butadiene monomer is recovered by steam-stripping and recycled.

STEP 2—Polybutadiene rubber is further polymerized, but in the presence of styrene and acrylonitrile monomers. This is done in low-pressure reactors under a nitrogen atmosphere. In this operation, the monomers are grafted onto the rubber backbone through the residual unsaturation remaining from the first step.

STEP 3—In a separate step, styrene-acrylonitrile (SAN) resin is prepared by emulsion, suspension, or mass polymerization by free-radical techniques. The operation is carried out in stainless-steel reactors operated at about 75°C (167°F) and 5 psig for about 7 hours. The final chemical operation is the blending of the ABS graft phase with the SAN resin, plus adding various antioxidants, lubricants, stabilizers, and pigments. Final operations involve preparation of a slurry of fine resin particles (via chemical flocculation), filtering, and drying in a standard fluid-bed dryer at 121–132°C (250–270°F) inlet air temperature.

It should be noted that many different polymerization techniques are used commercially to generate substrate, graft, and rigid backbone during the manufacture of ABS. These include emulsion, suspension, solution, and bulk polymerization. Also, twin-screw extruders, continuous mixers, and intensive internal fluxing mixers are used in addition to single-screw extruders as compounding devices.

Properties of ABS Resins. These resins incorporate a desirable balance of impact resistance, hardness, chemical resistance, a range of surface gloss levels, and non-staining properties. The dimensional stability of ABS is good; creep resistance is excellent. The resins have relatively low water absorption or change in volume over a range of humidities. Commercial forms of ABS include custom color-matched compounded pellets, natural pellets (unpigmented), and granular resins for use in compounding or alloying with other plastics.

Conversion of ABS Resins. A wide range of standard converting processes is applicable to ABS resins, including injection molding; compression molding; sheet extrusion followed by either pressure forming or vacuum forming; pipe extrusion; profile extrusion; and blow molding. ABS is slightly hygroscopic and should be dried prior to processing. Common fabricating techniques, such as drilling, sawing, riveting, solvent or ultrasonic welding, and insertion of metal fasteners can be used with ABS. Also, ABS can be hot stamped, painted, laminated, embossed, and, depending upon the grade of resin, electroplated or vacuum metallized.

Typical ABS Products. Injected molded parts can be found in telephone sets, refrigerator parts, plumbing fixtures and fittings, radios, televisions, small household appliances, business machines and personal computers, and numerous products for the automotive industry, including interior/exterior trim and bumpers. Examples of ABS extrusion grade product uses include refrigerator door and tank liners, bathtub surrounds, and boat hulls. These parts are made by extruding ABS into sheets and subsequently vacuum forming or pressure forming the sheet into the desired shape. ABS can also be directly extruded into the desired part, as with the extrusion of drain waste and vent (DWV) piping and profile extrusion of refrigerator breaker strips. Blow molding is an extrusion grade processing approach that is gaining in popularity and involves applications such as water tanks, interior and exterior doors, and some automotive parts. For these various applications, either general-purpose or specialty grades of ABS may be used. Some of the special properties that can be imparted to ABS follow:

Flame retardancy (FR) is a characteristic frequently needed in the manufacture of business machine and personal computer housings, some small appliances, and telecommunications equipment, among others. Flame retardancy can be imparted to ABS either by alloying with an inherently flame-retardant polymer, such as PVC (polyvinyl chloride), or by the use of a halogenated (predominantly bromine-related compounds) additive system. These FR grades of ABS can be designed to meet various flammability standards, such as those of Underwriters Laboratories, depending upon the needs of the application. In addition to flame retardancy, some business machine applications require RFI/EMI (radio frequency/electromagnetic interference) shielding. This is accomplished with ABS by giving the surface of the material a conductive coating via processes such as painting, electroplating, or vacuum metallizing.

High temperature resistance is a property of particular importance in the automotive interior trim area. Elevated temperature resistance is accomplished by either incorporating alpha methyl styrene or maleic anhydride in the basic formula, or by alloying ABS with polycarbonate. Heat deflection temperatures of 240°F (116°C) are possible by using these approaches.

Transparency can be imparted to ABS by using methyl methacrylate as the fourth monomer. The major use of transparent ABS is in the manufacture of refrigerator crisper trays.

Foamed ABS. The resin can be foamed either through the use of a chemical blowing agent, or by direct injection of nitrogen into the melt during processing. Structural foam grades of ABS are available in either general-purpose or flame-retardant formulas. Foam core pipe is manufactured by coextruding solid inner and outer skins around a structural foam wall interior. This is a widely practiced technique used by pipe extruders to reduce DWV pipe specific gravity, while not compromising pipe performance characteristics.

Other special characteristics can be imparted to ABS by various formula modifications. Some grades with specific thermal elongation characteristics can be electroplated or sputter coated. These materials find wide application in the automotive trim market. Gloss variations from matte-finish to high gloss are available to meet specific requirements. ABS can be reinforced with materials such as glass fibers, aluminum flake, mica, or barium ferrite. Although the ultraviolet (UV) stability of ABS has been historically poor, significant enhancement has recently been made by using special pigments, UV stabilizers, and paint systems. The use of polyelectrolytes in the formula has made it possible to offer grades of ABS with some antistatic qualities.

Engineering Staff, BorgWarner Chemicals,
Washington, West Virginia.

ABSAROKITE. A geologic term proposed by Iddings in 1805 for a porphyritic basalt containing phenocrysts of olivine and augite in a ground mass of smaller labradorite crystals. Type locality, Absaroka Range, Wyoming.

ABSCESS. A localized collection of pus within a cavity. An abscess may occur in many organs of the body. Abscesses can present a range of symptoms, depending upon location and cause, varying from severe, acute forms to milder, chronic forms. The presence of an abscess is always considered a serious matter by the physician because without immediate treatment, very serious consequences may occur.

Peritonsillar abscess (quinsy) develops between the capsule of the tonsil and the muscular bed of its fossa, being usually situated above and lateral to the tonsil. Such abscesses should be opened immediately and the cavity subjected to hot mouth washes while the patient is placed under standard antibiotic therapy.

Lung abscesses are among the more serious types. With the advent of antibiotics, the occurrence and severity of lung abscesses decreased markedly. Over half of the lung abscesses seen originate from a necrotizing suppurative bronchopneumonia resulting from the aspiration of mixed bacteria from the mouth and throat. Lung abscesses are sometimes associated with periodontal disease. In most cases, mixed anaerobic bacteria, such as *Fusobacterium nucleatum, Bacteroides melaninogenicus*, and anaerobic or microaerophilic streptococci and *Peptostreptoccus* predominate. *B. Fragilis* also may be present. Abscess formation is uncommon in pneumococcal pneumonia. Lung abscesses also may result from tumors or foreign bodies that cause bronchial obstruction. Other possible causes include pulmonary tuberculosis, fungal infection, and actinomycosis. Cough is present in nearly all patients. Copious foul sputum may result from drainage of the abscess into the bronchial tree. Chest pain and fever are common symptoms. X-rays may be required to confirm the presence of a lung abscess. Sputum examination is critical to diagnosis. Bronchoscopy is usually reserved for more difficult cases. Intravenous penicillin followed by oral penicillin V may be indicated. For penicillin-allergic patients, clindamycin may be used. Adequate drainage is also an important element of therapy. Where empyema is a complication of lung abscess, external drainage is required. Surgery is required in only a minimum of cases of lung abscesses, but was frequently required prior to antibiotics.

Liver abscess is most commonly caused by an amoeba (*Entameba histolytica*) which reaches the liver via the portal vein from sites where trophozoites have penetrated the colonic epithelium and the vessels of the submucosa. The abscess contents are not classically pus because they do not contain inflammatory cells and bacterial residues; the contents are usually sterile and contain only lysed liver cells. The patient has usually resided in the tropics and presents with pain in the right hypochondrium, epigastrium, and right shoulder. Some hepatomegaly is occasionally seen and, when the left lobe of the liver is affected, it may rupture into the pericardium, leading to signs of pericardial tamponade.

Metronidazole is the drug of first choice in treatment together with aspiration of the abscess contents, although in small abscesses the chemotherapy may suffice.

Intra-abdominal abscesses usually contain multiple bacterial species. Anaerobic bacteria are present in 60–70% of cases because of the proximity of the peritoneum to the bowel. The abscesses range from a small, acute disorder to a chronic process that causes intermittent fever, weight loss, and anemia. These "smouldering" abscesses sometimes result from a prior infection that was not fully eradicated by antibiotic therapy, allowing a pocket of infection to persist and slowly develop. Physicians, when attempting to diagnose a fever of undetermined origin, will usually ask the patient if there has been relatively recent abdominal surgery. The principal treatment for subphrenic abscesses is surgical drainage. Multiple antibiotic therapy is also used to reach a spectrum of possible causative organisms.

Although appendicitis is not classically considered an abscess, its treatment follows the general approach since delay in treatment can lead to generalized peritonitis or development of an abscess in the appendicular region.

Mediastinitis is an inflammation of the wall dividing the two pleural cavities; a common complication is mediastinal abscess. Sometimes the abscess opens and empties its contents into the trachea; either the patient will cough up large amounts of pus, or he may suffocate. Mediastinitis may occur as a result of perforation of the esophagus. This can happen when a sharp foreign body becomes lodged in the esophagus, during attempts to remove it, or during examination of the organ for other reasons. Mediastinitis also can result from a bullet or stab wound.

Abscess of the breast may occur within the first month after childbirth. It is caused by infections entering through a "cracked nipple." The unfortunate consequences of a breast abscess are that the infant is deprived of breast milk, plus the fact that the mother has a long period of discomfort and pain. Treatment is instituted as quickly as possible in order to avoid a prolonged convalescent period, as well as the possibility of the destruction of a large amount of breast tissue.

Anal infections may cause anal fissure, hemorrhoids, abscess, and fistula, and are usually the result of invasion of the numerous tiny glands or crypts, which abound in the tissues adjacent to the anus. If the infection spreads through the wall of the anus, an abscess may occur in the tissues around the anus, and this may burst through the skin around the anus or back into the rectum. In either case, the abscess cavity has two openings, the original site of entry of the infection and the point where it bursts through. Fistula is the term by which such a condition is designated.

In *abscess of the external ear*, there is pain and tenderness over the affected area. The auricle may enlarge to two or three times the normal size. If proper care is not given, the ear may be permanently distorted in shape. Antibiotics and sulfonamide drugs may be used effectively. Surgical treatment may be required, but only after careful examination by a specialist.

Intracranial abscesses are usually derived from middle ear infections, which account for two-thirds of the cases, while frontal sinus infections account for a further one-sixth. The infection is blood-borne after clinical septicemia and may follow tooth extraction. Although the infection may initially be entirely extradural, it will eventually penetrate the dura, causing a variety of problems—even ventriculitis. Cranial abscesses are usually surrounded by a considerable volume of edematous brain tissue and the mass effect of the edema can equal or exceed the potential damage of the abscess mass itself.

Epidural abscesses of the spinal cord are usually the result of *Staphylococcus* infections and are usually metastatic from other sites. They can, however, lead to a rapidly advancing paraplegia, and in such cases only myelography can distinguish the conditions. Treatment requires prompt evacuation and antibiotic therapy, but even then prognosis is not good because permanent damage may result to the spinal cord.

Periapical abscesses occur at the apical (apex) region of a tooth as the result of death of the pulp tissue. Periodontal abscesses occur in the tissues closely surrounding a tooth, such as gingiva, bone, or the periodontal membrane. When an abscess breaks through a limiting membrane, working through surrounding bone to external soft tissue, a gum boil may result.

R. C. Vickery, M.D.; D.Sc.; Ph.D., Blanton/Dade City, Florida.

ABSCISSION. This term is applied to the process whereby leaves, leaflets, fruits, or other plant parts become detached from the plant. Leaf abscission is a characteristic phenomenon of many species of woody dicots and is especially conspicuous during the autumn period of leaf fall. The onset of abscission seems to be regulated by plant hormones. Three main stages can be distinguished in the usual process of leaf abscission. The first is the formation of an abscission layer which is typically a transverse zone of parenchymatous cells located at the base of the petiole. The cells of this layer may become differentiated weeks or even months before abscission actually occurs. The second step is the abscission process proper which occurs as a result of a dissolution of the middle lamellae of the cells of the abscission layer. This results in the leaf remaining attached to the stem only by the vascular elements which are soon broken by the pressure of wind or the pull of gravity and the leaf falls from the plant. In the final stage of the process the exposed cells of the leaf scar are rendered impervious to water by lignification and suberization of the walls.

Subsequently other layers of corky cells develop beneath the outer layer. These layers eventually become a part of the periderm of the stem. The broken xylem elements of the leaf scar become plugged with gums or tyloses and the phloem elements become compressed and sealed off.

In some kinds of plants an abscission layer is only imperfectly formed and in many others, especially herbaceous species, no abscission layer develops at the base of the petiole. In a few herbaceous species, of which coleus, begonia, and fuchsia are examples, an abscission layer develops. In the majority of herbaceous species, however, and in some woody species, there is no true abscission process. In such herbaceous plants most or all of the leaves are retained until the death of the plant. In the woody plants falling in this category (example: shingle oak, *Quercus imbricaria*) the leaves are shed only by mechanical disruption from the plant. Abscission of the fruits of apple and doubtless of many other species occurs in much the same manner as abscission of leaves. The abscission of apple fruits can be artificially retarded by spraying with certain growth regulators.

Various plant hormones and plant growth regulators can be of help to the fruit producer in terms of controlling the timing of abscission. See also **Gibberellic Acid and Gibberellin Plant Growth Hormones;** and **Plant Growth Modification and Regulation.** See also related entries under **Tree.**

Abscission is covered in detail in "Abscission," by F. T. Addicott, University of California Press, Berkeley, California, 1982.

ABSOLUTE SPACE-TIME. A fundamental concept underlying Newtonian mechanics is that there exists a preferred reference system to which all measurements should be referred. This is known as absolute space-time. The assumption of such a system is replaced in relativistic mechanics by the principle of equivalence. See **Equivalence Principle; Relativity** and **Relativity Theory.**

ABSOLUTE TENSOR (Tensor Field). Tensor (tensor field) of weight zero. Often called tensor (tensor field) when context admits no confusion. See also **Tensor Field.**

ABSOLUTE ZERO. Conceptually that temperature where there is no molecular motion, no heat. On the Celsius scale, absolute zero is −273.15°C; on the Fahrenheit scale, −459.67°F; and zero degrees Kelvin (0K). The concept of absolute zero stems from thermodynamic postulations.

Heat and temperature were poorly understood prior to Carnot's analysis of heat engines in 1824. The Carnot cycle became the conceptual foundation for the definition of temperature. This led to the somewhat later work of Lord Kelvin, who proposed the Kelvin scale based upon a consideration of the second law of thermodynamics. This leads to a temperature at which all the thermal motion of the atoms stops. By using this as the zero point or absolute zero and another reference point to determine the size of the degrees, a scale can be defined. The Comité Consultative of the International Committee of Weights and Measures selected 273.16K as the value for the triple point for water. This set the ice-point at 273.15K.

From the standpoint of thermodynamics, the thermal efficiency E of an engine is equal to the work W derived from the engine divided by the heat supplied to the engine, $Q2$. If $Q1$ is the heat exhausted from the engine,

$$E = (W/Q2) = (Q2 - Q1)/Q2 = 1 - (Q1/Q2)$$

where W, $Q1$, and $Q2$ are all in the same units. A Carnot engine is a theoretical one in which all the heat is supplied at a single high temperature and the heat output is rejected at a single temperature. The cycle consists of two adiabatics and two isothermals. Here the ratio $Q1/Q2$ must depend only on the two temperatures and on nothing else. The Kelvin temperatures are then defined by the relation
where $Q1/Q2$ is the ratio of the heats rejected and absorbed, and $T1/T2$ is the ratio of the Kelvin temperatures of the reservoir and the source. If one starts with a given size for the degree, then the equation completely defines a thermodynamic temperature scale.

$$\frac{Q1}{Q2} = \frac{T1}{T2}$$

A series of Carnot engines can be postulated so that the first engine absorbs heat Q from a source, does work W, and rejects a smaller amount of heat at a lower temperature. The second engine absorbs all the heat rejected by the first one, does work and rejects a still smaller amount of heat which is absorbed by a third engine, and so on. The temperature at which each successive engine rejects its heat becomes smaller and smaller, and in the limit this becomes zero so that an engine is reached which rejects no heat at a temperature which is absolute zero. A reservoir at absolute zero cannot have heat rejected to it by a Carnot engine operating between a higher temperature reservoir and the one at absolute zero. This can be used as the definition of absolute zero. Absolute zero is then such a temperature that a reservoir at that temperature cannot have heat rejected to it by a Carnot engine which uses a heat source at some higher temperature.

ABSORBANCE. By combining the laws of Bouguer and Beer, the absorbance

$$A = -\log T = \log \frac{I_0}{I} = abc$$

where T is the transmittance, I_0 and I are the intensities of light incident and transmitted by a sample of thickness, b, concentration c (if the sample is in solution) and absorptivity a. It is assumed that all necessary corrections have been made in a reported value of A, hence terms such as absorbancy, absorptance, and absorptancy should not be used.

ABSORBER. In general, a medium, substance or functional part that takes up matter or energy. In radiation and particle physics, an absorber is a body of material introduced between a source of radiation and a detector to (1) determine the energy or nature of the radiation; (2) to shield the detector from the radiation; or (3) to transmit selectively one or more components of the radiation, so that the radiation undergoes a change in its energy spectrum. Such an absorber may function through a combination of processes of true absorption, scattering and slowing-down.

ABSORPTIMETRY. A method of instrumental analysis, frequently chemical, in which the absorption (or absence thereof) of selected electromagnetic radiation is a qualitative (and often quantitative) indication of the chemical composition of other characteristics of the material under observation. The type of radiation utilized in various absorption-type instruments ranges from radio and microwaves through infrared, visible, and ultraviolet radiation to x-rays and gamma rays. See also **Analysis (Chemical);** and **Spectro Instruments.**

ABSORPTION COEFFICIENT. 1. For the absorption of one substance or phase in another, as in the absorption of a gas in a liquid, the absorption coefficient is the volume of gas dissolved by a specified volume of solvent; thus a widely used coefficient is the quantity α in the expression $\alpha = V_0/V_p$, where V_0 is the volume of gas reduced to standard conditions, V is the volume of liquid and p is the partial pressure of the gas.

2. In the case of sound, the absorption coefficient (which is also called the acoustical absorptivity) is defined as the fraction of the incident sound energy absorbed by a surface or medium, the surface being considered part of an infinite area.

3. In the most general use of the term absorption coefficient, applied to electromagnetic radiation and atomic and subatomic particles, is a measure of the rate of decrease in intensity of a beam of photons or particles in its passage through a particular substance. One complication in the statement of the absorption coefficient arises from the cause of the decrease in intensity. When light, x-rays, or other electromagnetic radiation enters a body of matter, it experiences in general two types of attenuation. Part of it is subjected to scattering, being reflected in all directions, while another portion is absorbed by being converted into other forms of energy. The scattered radiation may still be effective in the same ways as the original, but the absorbed portion ceases to exist as radiation or is re-emitted as secondary radiation. Strictly, therefore, we have to distinguish the true absorption coefficient from the scattering coefficient; but for practical purposes it is sometimes convenient to add them together as the total attenuation or extinction coefficient.

If appropriate corrections are made for scattering and related effects, the ratio I/I_0 is given by the laws of Bouguer and Beer. Here, I_0 is the intensity or radiant power of the light incident on the sample and I is the intensity of the transmitted light. This ratio $I/I_0 = T$ is known as the transmittance. See also **Spectrochemical Analysis (Visible).**

ABSORPTION CURVE. The graphical relationship between thickness of absorbing material or concentration of dissolved substance and intensity of transmitted radiation.

ABSORPTION DISCONTINUITY. A discontinuity appearing in the absorption coefficient of a substance for a particular type of radiation when expressed as a function of the energy (or frequency or wavelength) of this radiation. An absorption discontinuity is often associated with anomalies in other variables such as the refractive index. See **Anomalous Dispersion.**

ABSORPTION EDGE. The wavelength corresponding to an abrupt discontinuity in the intensity of an absorption spectrum, notably an x-ray absorption spectrum, which gives the appearance of a sharp edge in the display of such a spectrum.

ABSORPTION (Energy). The process whereby the total number of particles emerging from a body of matter is reduced relative to the number entering as a result of interaction of the particles with the body. Also, the process whereby the kinetic energy of a particle is reduced while traversing a body of matter. This loss of kinetic energy or radiation is also referred to as moderation, slowing, or stopping. See also **Black Body.** The absorption of mechanical energy by dynamometers, which convert the mechanical energy to heat or electricity, has led to the use of the term "absorption dynamometer" to distinguish these machines. See also **Dynamometer**. In acoustics, absorption is the process whereby some or all of the energy of sound waves is transferred to a substance on which they are incident or which they traverse.

ABSORPTION (Physiology). The process by which materials enter the living substance of which the organism is composed. Materials including food and oxygen are taken into special organs by ingestion and respiration, but they must pass through the cell wall to become an integral part of the organism by absorption. The basic physical forces involved are those of osmosis and diffusion.

ABSORPTION (Process). Absorption is commonly used in the process industries for separating materials, notably a specific gas from a mixture of gases; and in the production of solutions such as hydrochloric and sulfuric acids. Absorption operations are very important to many air pollution abatement systems where it is desired to remove a noxious gas, such as sulfur dioxide or hydrogen sulfide, from an effluent gas prior to releasing the material to the atmosphere. The absorption medium is a liquid in which (1) the gas to be removed, i.e., absorbed is soluble in the liquid, or (2) a chemical reaction takes place between the gas and the absorbing liquid. In some instances a chemical reagent is added to the absorbing liquid to increase the ability of the solvent to absorb.

Wherever possible, it is desired to select an absorbing liquid that can be regenerated and thus recycled and used over and over. An example of absorption with chemical reaction is the absorption of carbon dioxide from a flue gas with aqueous sodium hydroxide. In this reaction, sodium carbonate is formed. This reaction is irreversible. However, continued absorption of the carbon dioxide with the sodium carbonate solution results in the formation of sodium acid carbonate. The latter can be decomposed upon heating to carbon dioxide, water, and sodium carbonate and thus the sodium carbonate can be recycled.

Types of equipment used for absorption include (1) a packed tower filled with packing material, absorbent liquid flowing down through the packing (designed to provide a maximum of contact surface), and gas flowing upward in a countercurrent fashion; (2) a spray tower in which the absorbing liquid is sprayed into essentially an empty tower with the gas flowing upward; (3) a tray tower containing bubble caps, sieve trays, or valve trays; (4) a falling-film absorber or wetted-wall column; and (5) stirred vessels. Packed towers are the most commonly used.

A representative packed-type absorption tower is shown in Fig. 1. In addition to absorption efficiency, a primary concern of the tower designer is that of minimizing the pressure drop through the tower. The principal elements of pressure drop are shown at the right of the diagram. Important to efficiency of absorption and pressure drop is the type of packing used. As shown by Fig. 2, over the years numerous types of packings (mostly ceramic) have been developed to meet a wide variety of operating parameters. A major objective is that of providing as much contact surface as is possible with a minimum of pressure drop. Where corrosion conditions permit, metal packing sometimes can be used. Of the packing designs illustrated, the berl saddles range in size from ¼ inch (6 millimeters) up to 2 inches (5 centimeters); raschig rings range from ¼ inch (6 millimeters) up to 4 inches (10 centimeters); lessing rings range from 1 inch (2.5 centimeters) up to 2 inches (5 centimeters); partition and spiral rings range from 3 inches (7.5 centimeters) up to 6 inches (15 centimeters).

In operation, the absorbing liquid is pumped into the top of the column where it is distributed by means of a weir to provide uniform dis-

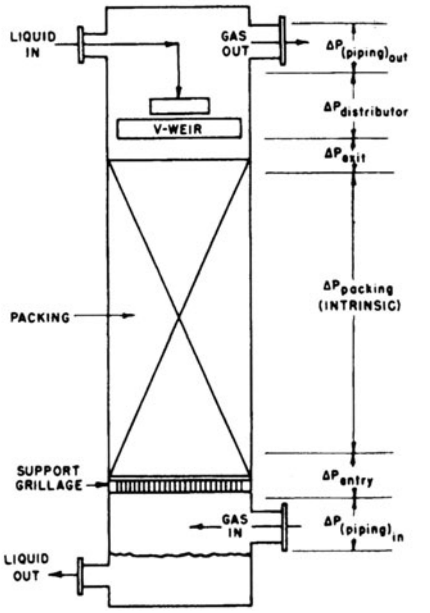


Fig. 1. Section of representative packed absorption tower.

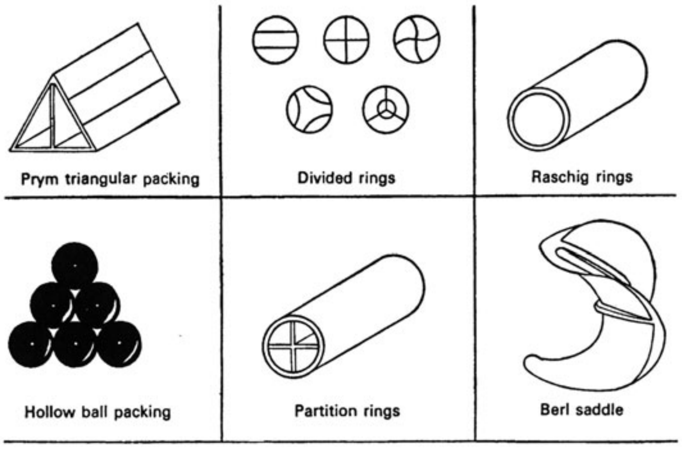


Fig. 2. Types of packing used in absorption towers.

tribution of the liquid over the underlying packing. Gas enters at the base of the tower and flows upward (countercurrent with the liquid) and out the top of the tower. The liquid may or may not be recycled without regeneration, depending upon the strength of the absorbent versus the quantity of material (concentration) in the gas to be removed. In a continuous operation, of course, a point is reached where fresh absorbing liquid must be added.

It is interesting to note that over 100,000 of the $\frac{1}{4}$-inch (6-millimeter) size packing shapes will be contained in each cubic foot (0.02832 cubic meter) of tower space if dense packing is desired.

In the purification of natural gas, the gas is fed into the bottom of an absorption tower where the gas is contacted countercurrently by a lean absorption oil. Hydrochloric acid is produced by absorbing gaseous hydrogen chloride in water, usually in a spray-type tower. Unreacted ammonia in the manufacture of hydrogen cyanide is absorbed in dilute sulfuric acid. In the production of nitric acid, ammonia is catalytically oxidized and the gaseous products are absorbed in water. The ethanolamines are widely used in scrubbing gases for removal of acid compounds. Hydrocarbon gases containing hydrogen sulfide can be scrubbed with monoethanolamine, which combines with it by salt formation and effectively removes it from the gas stream. In plants synthesizing ammonia, hydrogen and carbon dioxide are formed. The hydrogen can be obtained by countercurrently scrubbing the gas mixture in a packed or tray column with monoethanolamine which absorbs the carbon dioxide. The latter can be recovered by heating the monoethanolamine. In a nonliquid system, sulfur dioxide can be absorbed by dry cupric oxide on activated alumina, thus avoiding the disadvantages of a wet process. Sulfuric acid is produced by absorbing sulfur trioxide in weak acid or water.

See also **Coal; Ethanolamines; Chromatography;** and **Pollution (Air).**

ABSORPTION SPECTRUM. The spectrum of radiation which has been filtered through a material medium. When white light traverses a transparent medium, a certain portion of it is absorbed, the amount varying, in general, progressively with the frequency, of which the absorption coefficient is a function. Analysis of the transmitted light may, however, reveal that certain frequency ranges are absorbed to a degree out of all proportion to the adjacent regions; that is, with a distinct selectivity. These abnormally absorbed frequencies constitute, collectively, the "absorption spectrum" of the medium, and appear as dark lines or bands in the otherwise continous spectrum of the transmitted light. The phenomenon is not confined to the visible range, but may be found to extend throughout the spectrum from the far infrared to the extreme ultraviolet and into the x-ray region.

A study of such spectra shows that the lines or bands therein accurately coincide in frequency with certain lines or bands of the emission spectra of the same substances. This was formerly attributed to resonance of electronic vibrations, but is now more satisfactorily explained by quantum theory on the assumption that those quanta of the incident radiation which are absorbed are able to excite atoms or molecules of the medium to some (but not all) of the energy levels involved in the production of the complete emission spectrum.

A very familiar example is the spectrum of sunlight, which is crossed by innumerable dark lines—the Fraunhofer lines—from which so much has been learned about the constitution of the sun, stars, and other astronomical objects.

A noteworthy characteristic of selective absorption is found in the existence of certain anomalies in the refractive index in the neighborhood of absorption frequencies; discussed under **Dispersion (Radiation).** See also **Fraunhofer Lines.**

Additional Reading

Baeyans, W. R. G., et al.: "Luminescence Techniques in Chemical and Biochemical Analysis," in *Practical Spectroscopy Series*, Vol. 12, Marcel Dekker, New York, 1991.

Evans, N. J.: "Impedance Spectroscopy Reveals Materials Characteristics," *Adv. Mat. & Proc.*, 41 (November 1991).

Ewing, G. W., Editor: *Analytical Instrumentation Handbook*, Marcel Dekker, New York, 1990.

Grant, E. R., and R. G. Cooks: "Mass Spectrometery and Its Use in Tandem with Laser Spectroscopy," *Science*, 61 (October 5, 1990).

Margaritondo, G., Huber, D. L., and C. G. Olson: "Photoemission Spectroscopy of the High-Temperature Superconductivity Gap," *Science*, 770 (November 10, 1989).

Messerschmisr, E. F., and M. A. Harthcock: "Infrared Microspectroscopy: Theory and Applications," in *Practical Spectroscopy Series*, Vol. 8, Marcel Dekker, New York, 1988.

Miller, R. E.: "The Vibrational Spectroscopy and Dynamics of Weakly Bound Neutral Complexes," *Science*, 447 (April 22, 1988).

Robinson, J. W.: *Atomic Spectroscopy*, Marcel Dekker, New York, 1990.

Saykally, R. J.: "Infrared Laser Spectroscopy of Molecular Ions," *Science*, 157 (January 8, 1988).

Van Grieken, R., and A. Markowicz: *Handbook of X-Ray Spectrometry: Methods and Techniques*, Marcell Dekker, New York, 1992.

Various: "Application Reviews (Chemical Instrumentation)" *Analytical Chemistry* (*Special Issue*),(June 15, 1991).

Warren, W. S.: "Effects of Pulse Shaping in Laser Spectroscopy and Nuclear Magnetic Resonance," *Science*, 878 (November 11, 1988).

ABSORPTIVITY (Optical). If A is the absorbance of a solution b cm. in thickness and at a concentration c, the absorptivity is $a = A/bc$. See also **Beer's Law.**

ABUNDANCE RATIO. The proportions of the various isotopes making up a particular specimen of an element. See **Chemical Elements.**

ABYSSAL HILLS. Small hills averaging 100–200 meters (330–660 feet) in height that occupy the ocean floor. These may be nearly isolated or may occupy virtually the whole floor. See **Abyssal Plain.**

ABYSSAL PLAIN. An area on the ocean floor having a flat bottom and a very slight slope of less than 1 part in 1000. It is believed that these very flat surfaces arise from the continued deposition of mud and silt from turbidity currents. See **Ocean.** Seismographic studies support that these surfaces consist of such deposits. Mid-ocean canyons may be found on these abyssal plains; these are flat-bottomed depressions in the plains, varying from one to several kilometers (miles) in width and varying in depth up to several hundred meters (feet). These, too, are believed to be the product of certain turbidity currents.

ABYSSAL ROCKS. Proposed by Brögger as a general term for deep-seated igneous rocks, or those which have crystallized from magmas far below the surface of the earth, very slowly and under great pressure. Granite is a typical abyssal rock. The term plutonic is synonymous.

ABYSSAL ZONE. The region of the ocean beyond the point of penetration of light, including the ocean floor in the deep areas. According to various investigators who have descended into the ocean depths, no light penetrates beyond about 1,500 feet (450 meters), and penetration may be much less if the water is murky with suspended particles. The water is always extremely cold in the abyssal zone and the pressure is very great. Still many forms of animal life are to be found at these great depths, feeding upon the organic matter which drifts down from the upper waters.

Abyssal animals fall into two groups: *scavengers*, living on the shower of organic matter, and *predators*, which prey upon the scavengers or upon each other. The most abundant deep-sea animals are the sea cucumbers, snails, crustaceans, tunicates, cephalopods, and fish. The predaceous fish have large mouths filled with long, sharp teeth, and stomachs capable of great stretching; they are actually known to swallow fish larger than themselves. Many of the fish have a lure with a light on the end which attracts prey. Others have rows of light-producing organs on the sides of their bodies. Other animals with light-producing organs are coelenterates, echinoderms, annelids, crustaceans, and cephalopods. It is believed that the lights not only help these animals in finding food, but also in finding each other during the reproductive season. See also **Ecology; Ocean.**

ACACIA TREES. Of the family *Leguminosae* (pea family), the genus *Acacia* represents a large number of mostly evergreen trees and shrubs, particularly abundant in Africa and Australia. The trees like warmth and full sun. The small flowers are aggregated into ball-like or elongate clusters, which are quite conspicuous. The leaves are rather diverse in shape; quite commonly they are dissected into compound pinnate forms; in other instances, especially in the Australian species, they are reduced even to a point where only the flattened petiole, called a phyllode, remains. This petiole grows with the edges vertical, which some observers consider a protective adaptation against too intense sunlight on the surface. Some species, particularly those growing in Africa and tropical Asia, yield products of commercial value. Gum arabic is obtained from the *Acacia senegal*. A brown or black dye called clutch is obtained from *A. catechu*. Some acacias are used for timber. Shittinwood referred to in the scriptures: "And thou shalt make staves of shittinwood and overlay them with gold," (Exodus 26:26–37), is considered by authorities as wood from *Acacia seyal* (then referred to as the shittah tree).

Certain tropical American species are of particular interest because of the curious pairs of thorns, which are united at their base. These thorns are often hollowed out and used as nests by species of stinging ants. The leaves of some species, notably *Mimosa pudica*, are sensitive to the touch. The mimosa tree or silver wattle, native to Australia, is the *Acacia dealbata*. The leaves are fernlike and of a silver-green coloration. They attain a height of about 50 feet (15 meters) within 20 years, prefer full sun, and can be severely damaged by prolonged frosts. The tree has been introduced into warm regions of other parts of the world and has done well. The so-called catclaw acacia (*A. greggii*) has done well in the southwestern United States. One specimen, selected by The American Forestry Association for its "National Register of Big Trees," is located at Red Rock, New Mexico. The circumference at 4½ feet (1.4 meters) above the base is 6 feet, 5 inches (1.8 meters, 13 centimeters); the height is 49 feet (15 meters); and the spread is 46 feet (14 meters).

The koa acacia (*A. koa* var. hawaiiensis) is found in Hawaii. A specimen is also listed in the American Forestry Association registry and is located in the District of Kau. The circumference at 4½ feet (1.4 meters) above the base is 37 feet, 4 inches (11.4 meters); the height is 140 feet (42.7 meters); and the spread is 148 feet (45.1 meters).

A. baileyana, also a native of Australia, is known as the Cootamundra wattle or Bailey's mimosa. It attains a height of 20 feet (6 meters) or more, has long, narrow, waxy evergreen leaves of a silver-green color. The contour of the tree is often weeping.

The *Robinia pseudoacacia*, also referred to as black locust, common acacia, or false acacia, is found in the eastern United States. The tree is highly tolerant of dryness and industrial environments. This tree may attain a height of from 60 to 80 feet (18 to 24 meters), with a trunk diameter up to 4 feet (1.2 meters). It is a highly favored tree for gardens, often described as graceful and decorative.

Under normal environmental conditions, kudus (antelope family) that range freely in certain parts of South Africa (Transvaal) consume the leaves of acacia trees as a regular part of their diet. The leaves contain tannin, an astringent, at all times, but when the trees are under unusual stress, as may be caused by drought or overeating by herbivores, the tannin content of the leaves increases markedly. Even though this results in a very unpleasant taste, the kudus continue to eat the leaves to satiate their appetites during very dry seasons. Excessive consumption of the leaves causes inactivation of liver enzymes of the kudu, resulting in death. Researchers from the University of Pretoria and biologists from Dartmouth College have expanded on this situation, suggesting that other trees, such as sugar maple plants, also produce greater amounts of tannin and other defensive compounds when the leaves are damaged, as by herbivores. Another finding is that affected trees also emit ethylene gas, which can signal trees as far distant as 150 feet (50 meters) to activate the chemical defense process. Laboratory tests indicate that, when these plants are exposed to ethylene, tannin levels are increased in about 30 minutes. Researchers suggest that this may constitute a "natural population regulatory mechanism." Some ranchers who wish to maintain the kudu population now put out alfalfa during extremely dry periods. See also **Alleopathic Substance**.

Additional Reading

Staff: "The 1992 National Register of Big Trees," *Amer. Forests*, 14 (January–February 1992).

Yam, P.: "Acacia Trees kill antelope in the Transvaal," *Sci. Amer.*, 28 (December 1990).

ACANTHOCEPHALA (Thorny-Headed Worms). Worms, slender and hollow (pseudocoelom) with recurved hooks on invaginate proboscis, no digestive tracts, and adults parasitic in intestine of vertebrates, with larva in intermediate arthropod host. They are usually regarded as a class of roundworm (*Nematoda*), but ranking as a separate phylum is now favored.

ACANTHUS. Genus of the family *Acanthaceae* (acanthus family). This is a relatively small genus of Mediterranean plants grown mainly for ornamental purposes. The flowers are white or various shades of red. The term *ancanthus* also is used in architecture with reference to an ornamental design patterned after the leaves of the acanthus.

ACARICIDE. A substance, natural or synthetic, used to destroy or control infestations of the animals making up *Arachnida, Acarina*, mainly mites and ticks, some forms of which are very injurious to both plants and livestock, including poultry. There are numerous substances that are effective both as acaricides and insecticides; others of a more narrow spectrum are strictly acaricides. See also **Insecticide;** and **Insecticide and Pesticide Technology.**

ACARINA. The order of *Arachnida* which includes the mites and ticks.

ACCELERATED FLIGHT (Airplane). When the velocity of an airplane along its flight path contains elements of acceleration, the structure receives increments of inertial or dynamic loading that may prove to be far more severe upon the structure than the loading imposed by the static weight of the airplane and its contents. Consequently, accelerated flight has been the subject of extensive analytical and experimental investigation. Acceleration of rectilinear velocity below the speed of sound, as by increasing the thrust of the power plant in straight level flight, is of small import, since radial accelerations resulting from curvilinear flight at constant speed are so large as to be the critical influence. Cases of curved flight paths capable of accelerations of several *g* (acceleration due to gravity) are quick pull-ups (or "zooms") from high-speed rectilinear flight, spins, steeply banked turns, and loops. The magnitude of the effect of accelerated flight is well illustrated by considering the centrifugal force on an airplane following a curved flight path in the vertical plane. With a constant tangential speed as low as 120 mph (193 kph), the airplane experiences a radial acceleration of 4 *g* (4 times the acceleration of gravity) even though the radius of curvature be about 240 feet (73 meters).

An acceleration also is caused when the airplane encounters a gust. See also **Gust Front**.

ACCELERATION. The rate of change of the velocity with respect to the time is called acceleration. It is expressed mathematically by $d\mathbf{v}/dt$, the vector derivative of the velocity, $\mathbf{v}$ with respect to the time, t. If the motion is in a straight line whose position is clearly understood, it is convenient to treat the velocity v and the acceleration dv/dt as scalars with appropriate algebraic signs; otherwise they must be treated by vector methods.

Acceleration may be rectilinear or curvilinear, depending upon whether the path of motion is a straight line or a curved line. A body which moves along a curved path has acceleration components at every point. One component is in the direction of the tangent to the curve and is equal to the rate of change of the speed at the point. For uniform circular motion this component is zero. The second component is normal to the tangent and is equal to the square of the tangential speed divided by the radius of curvature at the point. This normal component, which is directed toward the center of curvature, also equals the square of the angular velocity multiplied by the radius of curvature. The acceleration due to gravity is equal to an increase in the velocity of about 32.2 feet (981.5 centimeters)/second/second at the earth's surface and is of prime importance since it is the ratio of the weight to the mass of a body. For examples of acceleration in both curved and linear motion, see **Kinematics**. See also **Angular Velocity** and **Angular Acceleration**.

ACCELERATION (Due to Gravity). The universal character of the gravitational force for point masses or spherical bodies can be expressed by the equation:

$$F = \frac{GM_1M_2}{R^2}, \tag{1}$$

where

The constant G is independent of all properties of the particular bodies

M_1, M_2 = masses of two bodies

R = distance between two bodies

G = a constant = 6.670×10^{-8} dyne cm^2 gm^{-2}

involved.

The weight of a body of mass M on the earth is the force with which it is attracted to the center of the earth. On the surface of the earth, the weight is given by:

$$W = Mg, \tag{2}$$

where the acceleration due to gravity is obtained from Equation (1):

$$\begin{aligned} g &= \frac{GM_E}{R_E^2} \\ &= 980.665\ \text{cm/(second)}^2 \\ &= 32.174\ \text{feet/(second)}^2\ (9.81\ \text{meters/(second)}^2) \end{aligned} \tag{3}$$

Variation of the acceleration due to gravity at sea level for different latitudes on earth is given in the accompanying table.

VARIATION OF ACCELERATION DUE TO GRAVITY ON EARTH WITH LATITUDE (At Sea Level)

Latitude	g centimeters/(second)^2	g feet/(second)^2
0°	977.989	32.0862
10°	978.147	32.0916
20°	978.600	32.1062
30°	979.295	32.1290
40°	980.147	32.1570
50°	981.053	32.1867
60°	981.905	32.2147
70°	982.600	32.2375
80°	983.053	32.2523
90°	983.210	32.2575

Gravitational force at the surface of the sun, moon, and planets is given in specific entries on these solar system bodies. See also **Gravitation.**

ACCELERATOR (Particle). See **Particles (Subatomic).**

ACCELEROMETER. Acceleration is not measured directly, but is computed by measuring the force exerted by restraints that are placed on a mass to hold its position fixed in an accelerating body. The relationship between restraint and acceleration is defined by Newton's second law: $F = ma$ (force equals mass times acceleration). The device used to accomplish the foregoing requirement is known as an *accelerometer*. The accelerometer is the most common type of vibration sensor. See also **Vibration.** The necessary attributes of an accelerometer include small size, light weight, and rugged construction. Both self-generating accelerometers and those requiring electrical excitation are available. The most common is the self-generating piezoelectric device. Typical performance characteristics for accelerometers are given in the accompanying table.

Vibration sensing is required over a wide range of amplitudes and frequencies. For example, vibration from rotating machinery occurs from about 1 Hz to over 20,000 Hz, with the principal interest being from 10 to 2000 Hz. A smooth running motor may vibrate at .01g (1g = 386 in./sec^2; 980 cm/sec^2), but a high-speed gearbox can easily vibrate at more than 100g at a frequency of over 10,000 Hz. Ideally, a vibration sensor attaches to a body in motion and provides an output signal proportional to the vibrational input from that body. See Fig. 1. Sometimes it is not practical to attach a sensor directly to the moving body. In those cases, the measurement is made by attaching the sensor to another body and making a measurement relative to the motion of that body. In any case, measurement performance must not be degraded because of the location of the sensor.

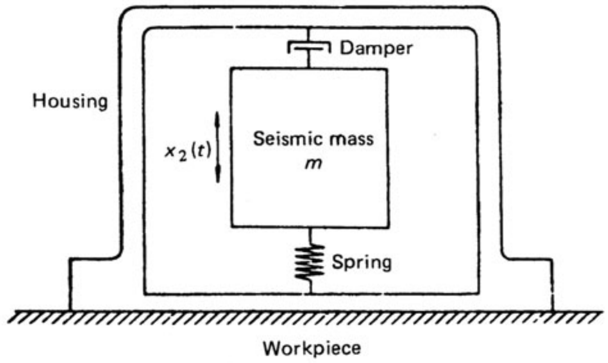


Fig. 1. Mechanical schematic of inertial sensor.

Piezoelectric Accelerometers. These devices utilize a mass in direct contact with a piezoelectric component, or crystal. When a varying motion is applied to the accelerometer, the crystal experiences a varying force excitation ($F = ma$) causing a proportional electrical charge q to be developed across it:

$$q = d_{ij}F = d_{ij}ma$$

where d_{ij} is the material's piezoelectric strain constant.

As the equation shows, the electrical output from the piezoelectric material is dependent on its properties. Two commonly used materials are lead zirconate titanate ceramic (PZT) and quartz. As self-generating materials, they both produce a large electrical charge for their size, although the piezoelectric strain constant of PZT is about 150 times that of quartz. As a result, accelerometers using PZT are more sensitive or are much smaller. The mechanical spring constants for the piezoelectric components are high and the inertial masses attached to them are small. Therefore, these accelerometers are useful to extremely high frequencies. Damping is rarely added to these devices. Fig. 2 shows a typical frequency response for such a device. Piezoelectric accelerometers have comparatively low mechanical impedance. Therefore, their effect on the motion of most structures is small. They are also rugged and have stable outputs with time and environment.

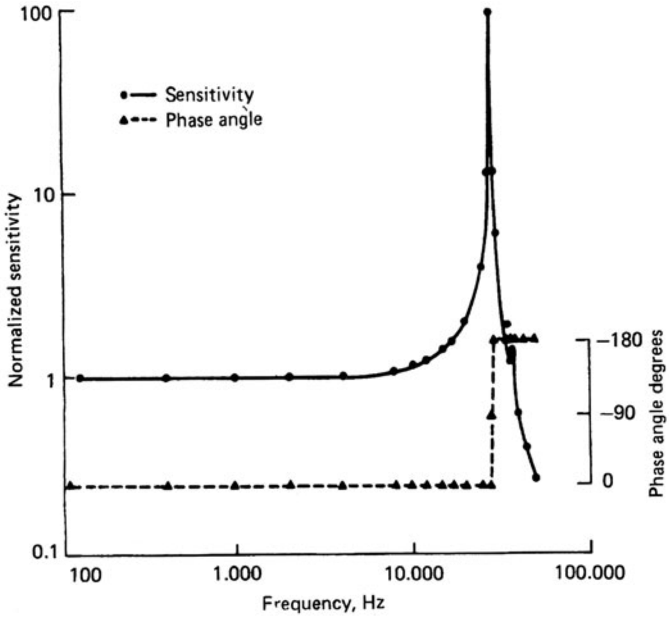
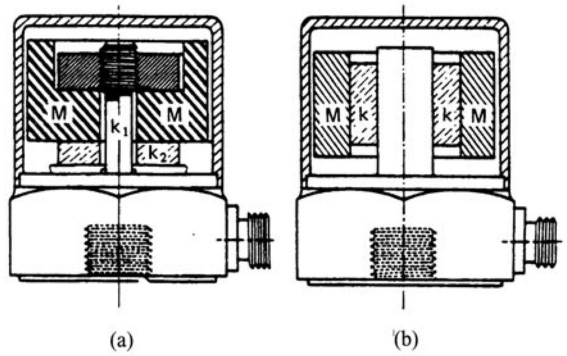


Fig. 2. Frequency response from typical piezoelectric accelerometer.

Two principal design configurations are used for piezoelectric accelerometers. One stresses the piezoelectric material in *compression* while the other stresses it in *shear*. Simple diagrams of these are illustrated by Fig. 3. When the accelerometer is accelerated upward, the mass is moved downward toward the bottom of the accelerometer. Conversely, downward acceleration moves the mass element upward. With vibration motion, the resultant dynamic stress deforms the piezoelectric element. For the compression accelerometer, vibration varies the stress in the crystal which is held in compression by the pre-load element. For the shear accelerometer vibration simply deforms the crystal in shear. The mechanical construction for actual designs can be more complex, but the model is the same.

Piezoresistive Accelerometers and Strain-Gage Sensors. Piezoresistive accelerometers are strain gage sensors which use semiconductor strain gages in order to provide much greater gage factors than possible with metallic gages. Higher gage factors are achieved because the material resistivity changes with stress, not just its dimensions. The increased sensitivity is critical to vibration measurement in that it permits miniaturization of the accelerometer.



(a) (b)

Fig. 3. Conceptual diagram for piezoelectric accelerometers: (a) compression, and (b) shear.

The typical piezoresistive accelerometer uses either two or four active gages in a Wheatstone bridge. It is more important to use multiple gages than when using metallic gages because the temperature coefficients of the semiconductor elements are greater than metallic gages. To control the electrical bridge balance and sensitivity variations with temperature, other resistors are used within the bridge and in series with the input.

The mechanical construction of an inertial system using piezoresistive elements is illustrated by Fig. 4. The construction of a complete accelerometer is shown by Fig. 5. This design includes overload stops to protect the gages from high amplitude inputs and includes oil to improve damping. Such an instrument is useful for acquiring vibration information at low frequencies (for example, below 1 Hz) and the device can be used to sense static acceleration.

Servo Accelerometers. The construction and operating principle of a servo accelerometer are illustrated by Fig. 6. When subjected to acceleration, the proof mass deflects relative to the base of the accelerometer, and the pickoff changes its capacitance as a result of changes in the damping gap. As this occurs, the servo supplies current to the coil which is located in the gaps of the permanent magnets. The resulting force restores the coil to its equilibrium position. The output signal is a measure of the coil current and is proportional to the applied acceleration.

Signal Conditioning for Accelerometers. Signal conditioners interface accelerometers to readout and processing instruments by (1) providing power to the accelerometer, if it is not self-generating, (2) providing proper electrical load to the accelerometer, (3) amplifying the signal, and (4) providing appropriate filtering and drive signal. Piezoelectric and piezoresistive transducers both require conditioners with certain characteristics as is now discussed.

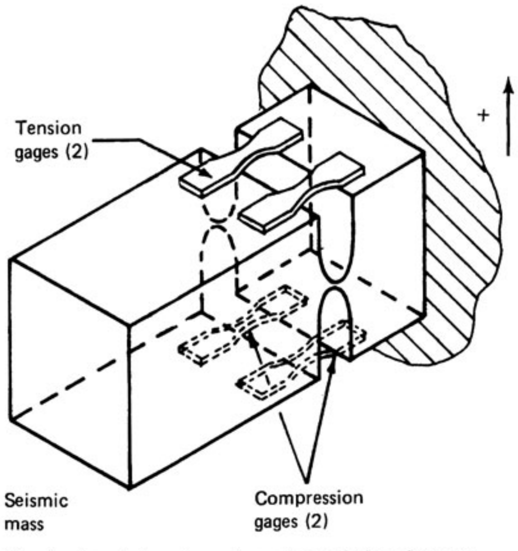


Fig. 4. Inertial system using piezoresistive elements.

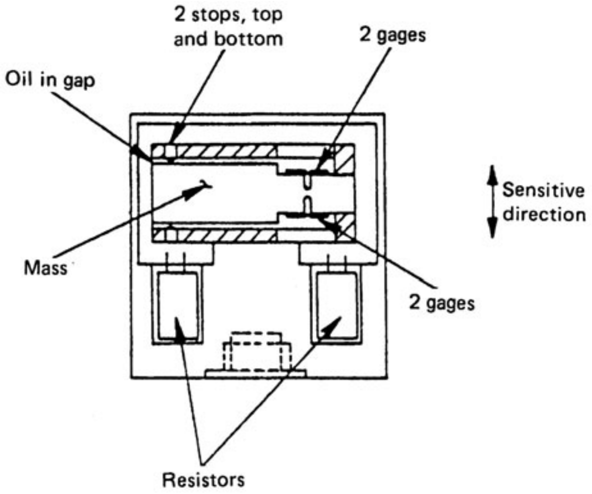


Fig. 5. Construction of typical piezoresistive elements.

Conditioning Piezoelectric Accelerometers. The piezoelectric accelerometer is self-generating and supplies a very small amount of energy to the signal conditioner. It presents a very high source impedance, mainly capacitive, to the conditioner. Two forms of schematic representation of a piezoelectric accelerometer are shown in Fig. 7. It may be regarded as a voltage source in series with a capacitance, or as a charge source in parallel with a capacitance. The signal conditioner determines how the transducer is treated in a given system. Both voltage and charge sensing are used. The charge amplifier is by far the most common approach. The charge amplifier is advantageous because the system gain and low frequency response are well defined and are independent of the cable length and accelerometer capacitance.

The charge amplifier consists of a charge converter and a voltage amplifier as shown in Fig. 8. The system does not amplify charge per se. It converts input charge to a voltage and then amplifies the voltage.

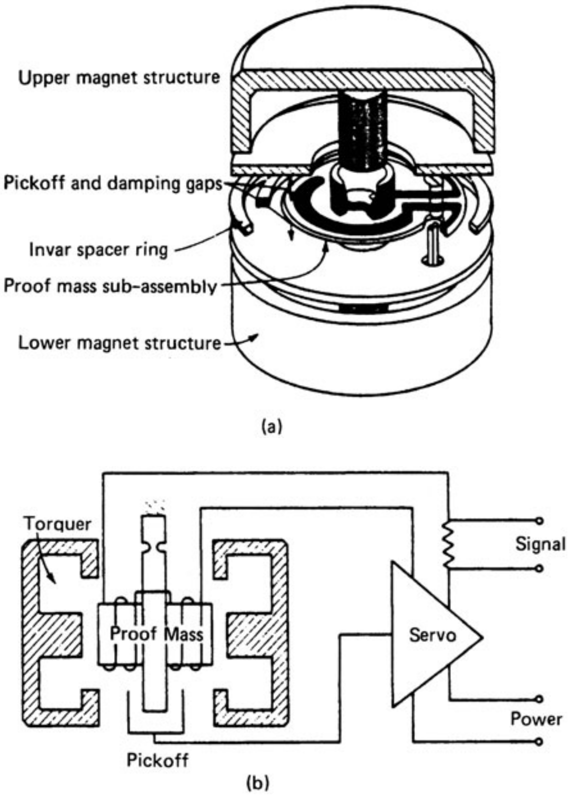


Fig. 6. Construction of serve accelerometer.

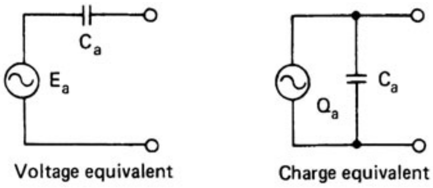


Fig. 7. Electrical schematic representation of piezoelectric accelerometers.

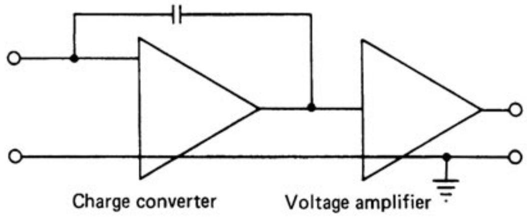


Fig. 8. Block diagram of charge amplifier.

A charge converter is essentially an operational amplifier with integrating feedback. The equivalent circuit is shown in Fig. 9. With basic operational type feedback, amplifier input is maintained at essentially zero volts and therefore looks like a short circuit to the input. Amplifier output is a function of input current. Having integrating operational feedback, output is the integral of input current. Hence the name "charge amplifier," since

$$q = \int i\, dt.$$

In operation, the charge converter output voltage which occurs as a result of a charge input signal, is returned through the feedback capacitor to maintain the voltage at the input at, or close to, zero. Thus, the charge input is stored in the feedback capacitor, producing a voltage across it which is equal to the value of the charge input divided by the capacitance of the feedback capacitor. The transfer characteristic (charge gain) of this amplifier is therefore solely dependent upon the value of the feedback capacitor (or network if a more complicated feedback is used) provided that the amplifier's open loop gain and input impedance are sufficiently high. The midband charge gain (*mV/pC*) of this circuit is

$$\frac{E_o}{Qi} = \frac{1}{C_f}.$$

The complete circuit transfer function is given by:

$$\frac{E_o}{Q_i} = \frac{-A}{1+A} \cdot \frac{sR_f}{sR_f\left\{\dfrac{C_s + C_c}{1+A}\right\} C_f + 1}$$

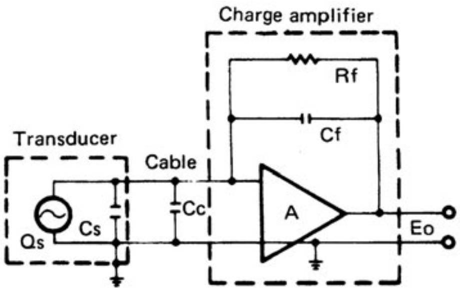


Fig. 9 Equivalent circuit of charge converter.

where: E_o = charge converter output
Q_t = accelerometer signal (charge)
s = accelerometer sensitivity (charge)
C_a = accelerometer capacitance
C_c = cable capacitance
C_f = feedback capacitance
R_f = feedback resistance
A = amplifier open looop gain

To simplify, we can assume:

$$\frac{-A}{1+A} = -1$$

since amplifier open loop gain A is very high and $A \approx 1 + A$.

Because C_f is very large compared to $(C_s + C_c)/(1 + A)$ in most applications, system gain is independent of cable length. Only an unlikely combination of extremely large source or cable capacitance and very small feedback capacitance would influence system gain, and then only to a minor degree. Therefore, charge measuring systems are seldom calibrated end-to-end.

With the above simplifications, the denominator of the expression becomes $sR_fC_f + 1$. This is the characteristic form of a simple first order rolloff at

$$f_{-3\text{dB}} = \frac{1}{2R_fC_f}$$

with a terminal slope approaching 6 dB per octave. For all practical purposes, low frequency response of a charge measuring system is a function of well defined electronic components and does not vary with cable length. This is a very important feature when measuring low frequency vibrations.

Conditioning for Low Impedance Piezoelectric Accelerometers. Piezoelectric accelerometers are available with simple electronic circuits internal to their cases to provide signal amplification and low impedance output. Some designs operate from low-current dc voltage supplies and are designed to be *intrinsically safe* when coupled by appropriate barrier circuits. Other designs have common power and signal lines and use coaxial cables.

The principal advantages of piezoelectric accelerometers with integral electronics are their relative immunity to cable-induced noise and spurious response, the ability to use lower cost cable, and lower signal conditioning cost. In the simplest case, the power supply might consist of a battery, a resistor, and a capacitor.

These advantages do not come without compromise. Because the impedance matching circuitry is built into the transducer, gain cannot be adjusted to utilize the wide dynamic range of the basic transducer. Ambient temperature is limited to that which the circuit will withstand, and this is considerably lower than that of the piezoelectric sensor itself. In order to retain the advantages of small size, the integral electronics must be kept relatively simple. This precludes the use of multiple filtering and dynamic overload protection and thus limits their application. But when conditions are relatively benign, these accelerometers can economically provide excellent noise immunity and signal fidelity.

Conditioning Piezoresistive Transducers. Piezoresistive transducers are relatively easy to condition. They generally have high level output, low output impedance, and very low intrinsic noise. These transducers require an external power supply. This supply is usually dc, but it may be ac providing the carrier frequency is at least five to ten times the maximum frequency of interest.

Most transducers are designed for constant-voltage excitation and are used with relatively short cables. With long cables, wire resistance is not negligible. Moreover, resistance will change with temperature, and the voltage drop along the line will vary as transducer resistance or load changes. For these applications, transducers should be calibrated for constant-current excitation so their output will be less dependent upon external effects.

Many piezoresistive transducers are full-bridge devices. Some have four active arms to maximize sensitivity. See Fig. 10. Others have two active arms and two fixed precision resistor arms to permit shunt calibration by precision calibration resistors in the signal conditioner. Miniature transducers are usually half-bridge devices, with bridge completion accomplished in the signal conditioner.

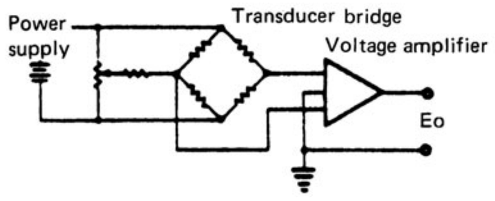


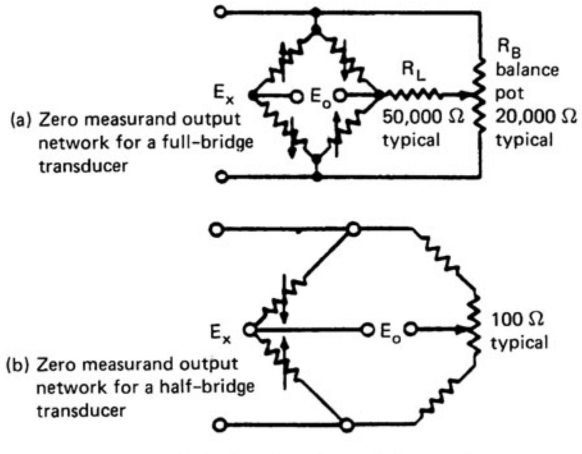
Fig. 10. Typical system and bridge circuit for piezoresistive accelerometer.

Adjustment of the unbalanced output of an accelerometer can easily be performed in the signal conditioner. For full-bridge transducers, the balancing potentiometer R_1 is connected across the excitation terminals and a current-limiting resistor is connected between the wiper arm of the potentiometer and the bridge. This is shown in Fig. 11(a). For half-bridge transducers, a small balance potentiometer (typically 100 ohms) is connected between the bridge-completion arms as shown in Fig. 11(b).

Environmental Effects. *Temperatures.* Accelerometers can be used over wide temperature ranges. Piezoelectric devices are available for use from cryogenic temperatures (−270°C) to over 650°C. Sensitivity changes with ambient temperature, but the changes are systematic and can be calibrated. If the ambient temperature changes suddenly so that strains develop within the accelerometer and within the time response of the measurement system, further errors can occur. These are evaluated by testing the response of accelerometers to step function changes of temperature per industry standard test procedures. Errors usually appear as a wandering signal, or a low frequency oscillation.

Cable Movement. Cabling from the accelerometer to the signal conditioner can generate spurious signals when it is subjected to movement or dynamic forces. This is usually only significant for systems using high impedance piezoelectric accelerometers. The major noise generating mechanism is triboelectric noise. This is caused by charge trapping due to relative motion, or localized separation between the cable dielectric and the outer shield around the dielectric. To reduce this effect, cabling is available which is "noise treated." These cables have a conductive coating applied to the surface of the dielectric which prevents charge trapping. Another method used to eliminate this effect is to use a sensor which includes an electronic circuit to reduce the impedance to about 100 ohms.

Dynamic Strain Inputs. In vibration environments, some structures may dynamically flex, stretch, or bend at the mounting location of the accelerometer. The base of the accelerometer being in intimate contact with this strained area, can also be strained. A portion of this base strain is transmitted directly to the crystal sensing element, and will generate error signals. In addition to strains in the structure, it is also possible to induce errors from forces or pressures onto the case of the accelerometer. Outputs from these vary greatly, depending upon the internal design



(a) Zero measurand output network for a full-bridge transducer

(b) Zero measurand output network for a half-bridge transducer

Fig. 11. Bridge balancing for piezoresistive accelerometer.

TYPICAL ACCELEROMETER PERFORMANCE CHARACTERISTICS

Characteristics	Piezoelectric Accelerometers		Piezoresistive Accelerometer	Servo Accelerometer
Sensitivity, pC/g, mV/g	10	100	20	250
Frequency range, Hz	4–15,000	1–5,000	0–750	0–500
Resonance frequency, Hz	80,000	20,000	2,500	1,000
Amplitude range, g	500	1,000	25	15
Shock rating, g	2,000	10,000	2,000	250
Temperature range, °C	−50 to +125	−50 to +260	0 to +95	−40 to +85
Total mass, g	1	29	28	80

of the accelerometer. The errors from these are usually checked to industry standard test procedures and the results are included in specifications.

Electrostatic and Electromagnetic Fields. Electrostatic noise can be generated by stray capacitance coupling into the measurement system. It is important that the cabling between a high impedance piezoelectric sensor and the signal conditioner be fully shielded. Ground loops can be avoided by grounding the system at a single point, usually at the output of the signal conditioner. Magnetically coupled noise can best be avoided by not placing signal cables in close proximity to power or high current conductors, and by avoiding electromagnetic sources when possible. Accelerometers should also be checked for their sensitivity to electromagnetic fields.

Velocity Transducers. Electrodynamic velocity pickups have the capability of being used directly into readout instruments. The self-generating voltage in the transducer is proportional to the velocity of the motion being measured and is usually of sufficient amplitude so that no voltage amplification is required. The disadvantages of velocity pickups are their large size and mass and their inability to be used for measurements at frequencies below about 10 Hz. Also, the output at high frequencies, above about 1000 Hz, is quite small in most applications. Care must be taken in using these devices in strong magnetic field environments.

A typical velocity sensor consists of a seismically mounted, and usually critically damped, magnetic core suspended in a housing which is rigidly attached to the vibrating surface. A coil of wire attached to the housing surrounds the core. Relative motion between the magnetic core and the housing causes magnetic lines of flux to cut the coil, inducing a voltage proportional to velocity. These sensors operate above their first natural frequency.

Relative Motion Sensing—Eddy Current Probe. In some cases, it is not practical to place a sensor in contact with the moving part. Relative motion measurement approaches are then used. The most commonly used device is the eddy current probe. Non-contact eddy current displacement measuring systems have achieved general acceptance for industrial machinery protection and condition monitoring. An eddy current displacement probe, generally about 0.300 in. (7.5 mm) in diameter, contains a small coil of fine wire at its tip which is excited by a remote radio frequency oscillator to generate a magnetic field. As the tip of the probe is brought close to a conductive surface, such as a rotating shaft, eddy currents induced in the conductor by the probe's magnetic field oppose the field and reduce the amplitude of the carrier by an amount proportional to the change in proximity. A demodulator, usually encapsulated in the same enclosure as the oscillator, converts the change in carrier amplitude to a low impedance, calibrated voltage output.

An eddy current displacement sensor and its companion oscillator-demodulator is therefore a gap-to-voltage measuring system. The average gap, or the distance between the probe tip and the conductive surface, is represented by a dc bias or offset on which is superimposed an ac analog of the surface's dynamic motion. Typical linear amplitude range is 1 to 2 mm with a frequency response capability of dc to more than 2000 Hz. Sensitivity changes for different target materials and with changes in cable length.

Additional Reading

Anon.: "ISA Recommended Practice for Specifications and Tests for Piezoelectric Acceleration Transducers (ISA 37.2)," Instrument Society of America, Research Triangle Park, North Carolina (updated periodically).

Anon.: "ANSI Standard for the Selection of Calibration and Tests for Electrical Transducers Used for Measuring Shock and Vibration (ANSI Rep. S2.11)," American National Standards Institute, New York (updated periodically).

Link, B., Suminto, J., and M. Young: "A Rugged, High-Performance Piezoresistive Accelerometer," *Sensors*, 43 (October 1991).

Meirovitch, L.: "Elements of Vibration Analysis," 2nd Edition, McGraw-Hill, New York, 1985.

Robert M. Whittier, Endevco Corp.,
San Juan Capistrano, California.

ACCOMMODATION COEFFICIENT. A quantity defined by the equation

$$a = \frac{T_3 - T_1}{T_2 - T_1}$$

where T_1 is the temperature of gas molecules striking a surface which is at temperature T_2, and T_3 is the temperature of the gas molecules as they leave the surface, a is the accommodation coefficient. It is, therefore, a measure of the extent to which the gas molecules leaving the surface are in thermal equilibrium with it.

ACCOMMODATION (Ocular). The mechanism whereby the equatorial diameter of the lens of the eye may be decreased and its thickness increased to focus clearly on the retina the image of a near object.

ACCRETION (Geology). The process by which crystals and other solid bodies grow by the addition of material onto their surfaces. A concretion is a body that grows from the center outward in a regular manner by successive additions of material.

ACCUMULATOR (Hydraulic). The hydraulic accumulator is a hydraulic device consisting of a cylinder and piston which is actuated by weight, springs, or compressed fluid. On the opposite side of the piston a fluid such as water, oil, air, etc., is stored and, consequently, is available to do work when the pressure on it is reduced. This work is obtained by virtue of the fact that while the fluid is discharged rapidly, giving large hydraulic power for short periods of time, it may be refilled by a comparatively small and low-powered pump working a much longer period of time.

One successful type of accumulator consists of a closed cylinder containing a rubber bag filled with air or gas. Fluid is pumped into the cylinder under pressure, compressing the air or gas in the bag. When the release valve is opened the air or gas in the bag expands to eject the fluid from the cylinder.

Another type of hydraulic accumulator is the pumped storage plant, used in electric power systems for the economic carrying of variable loads. As employed in conjunction with steam generating stations, steam turbine-driven centrifugal pumps raise water from a lower to an upper pool with off-peak power. During the peak-load periods this water is released to the lower pool through a hydraulic turbo-generator as rapidly as is needed to give the required power. In favorable locations the overall efficiency of conversion and storage need not to be greater than 50% in order to justify the project.

ACCURACY. In terms of instruments and scientific measuring systems, accuracy may be defined as the conformity of an indicated value to an accepted standard value, or true value. Accuracy is usually measured in terms of *inaccuracy* and expressed as *accuracy*. As a performance specification, accuracy should be assumed to mean *reference accuracy* unless otherwise stated.

Reference accuracy may be defined as a number or quantity which defines the limit that errors will not exceed when the device is used under reference operating conditions. See accompanying diagram. Reference accuracy includes the combined conformity, hysteresis, and repeatability errors. The units being used must be stated explicitly. It is preferred that a + and − sign precede the number or quantity. The absence of a sign infers both a plus and a minus sign. Reference accuracy can be expressed in a number of forms, of which the five following examples are typical:

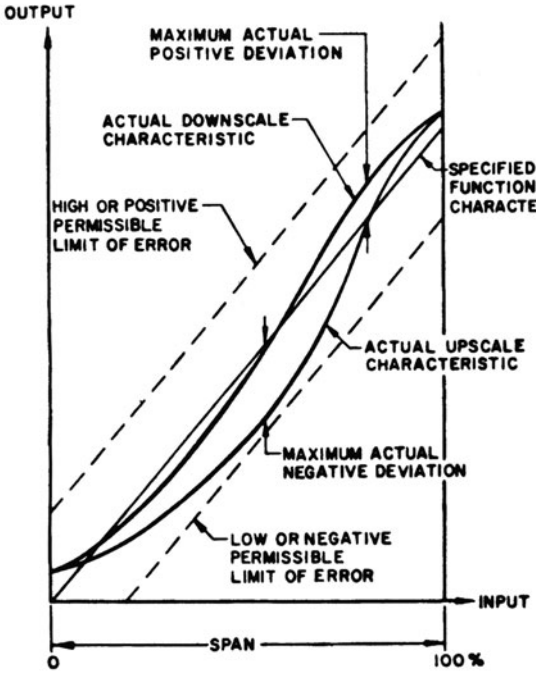
(1) Reference accuracy expressed in terms of the *measured variable*. Typical expression: The reference accuracy is ± 1°F (± 0.6°C).

(2) Reference accuracy expressed in percent of *span*. Typical expression: The reference accuracy is ± 1/2% of span. This percentage is calculated using units, such as deg. F or C, psi, and so on.

(3) Reference accuracy expressed in percent of the *upper range-value*. Typical expression: The reference accuracy is ± 1/2% of upper range-value. This percentage is also calculated, using units as previously given.

(4) Reference accuracy expressed in percent of *scale length*. Typical expression: The reference accuracy is ± 1/2% of scale length.

(5) Reference accuracy expressed in percent of *actual output reading*. Typical expression: The reference accuracy is ± 1% of actual output reading. Again, this percentage is calculated, using scale units such as previously given.

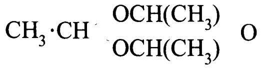


Fundamental relationships pertaining to reference accuracy.

ACETALDEHYDE. CH_3CHO, formula weight 44.05, colorless, odorous liquid, mp −123.5°C, bp 20.2°C, sp gr 0.783. Also known as *ethanal*, acetaldehyde is miscible with H_2O, alcohol, or ether in all proportions. Because of its versatile chemical reactivity, acetaldehyde is widely used as a commencing material in organic syntheses, including the production of resins, dyestuffs, and explosives. The compound also is used as a reducing agent, preservative, and as a medium for silvering mirrors. In resin manufacture, paraldehyde $(CH_3CHO)_3$ sometimes is preferred because of its higher boiling and flash points.

In tonnage production, acetaldehyde may be manufactured by (1) the direct oxidation of ethylene, requiring a catalytic solution of copper chloride plus small quantities of palladium chloride, (2) the oxidation of ethyl alcohol with sodium dichromate, and (3) the dry distillation of calcium acetate with calcium formate.

Acetaldehyde reacts with many chemicals in a marked manner, (1) with ammonio-silver nitrate ("Tollen's solution"), to form metallic silver, either as a black precipitate or as an adherent mirror film on glass, (2) with alkaline cupric solution ("Fehling's solution") to form cuprous oxide, red to yellow precipitate, (3) with rosaniline (fuchsine, magenta), which has been decolorized by sulfurous acid ("Schiff's solution"), the pink color of rosaniline is restored, (4) with NaOH, upon warming, a yellow to brown resin of unpleasant odor separates (this reaction is given by aldehydes immediately following acetaldehyde in the series, but not by formaldehyde, furfuraldehyde or benzaldehyde), (5) with anhydrous ammonia, to form aldehyde-ammonia $CH_3\cdot CHOH\cdot NH_2$, white solid, mp 97°C, bp 111°C, with decomposition, (6) with concentrated H_2SO_4, heat is evolved, and with rise of temperature, paraldehyde $(C_2H_4O)_3$ or



$$CH_3\cdot CH\begin{matrix}OCH(CH_3)\\OCH(CH_3)\end{matrix}O$$

colorless liquid bp 124°C, slightly soluble in H_2O, is formed, (7) with acids, below 0°C, forms metaldehyde $(C_2H_4O)_x$, white solid, sublimes at about 115°C without melting but with partial conversion to acetaldehyde, (8) with dilute HCl or dilute NaOH, aldol, $CH_3\cdot CHOH\cdot CH_2\cdot CHO$ slowly forms, (9) with phosphorus pentachloride, forms ethylidene chloride, $CH_3\cdot CHCl_2$, colorless liquid, bp 58°C, (10) with ethyl alcohol and dry hydrogen chloride, forms acetal, 1,1-diethyoxyethane $CH_3\cdot CH(OC_2H_5)_2$, colorless liquid, bp 104°C, (11) with hydrocyanic acid, forms acetaldehyde cyanhydrin, $CH_3\cdot CHOH\cdot CN$, readily converted into alphahydroxypropionic acid, $CH_3\cdot CHOH\cdot COOH$, (12) with sodium hydrogen sulfite, forms acetaldehyde sodium bisulfite, $CH_3\cdot CHOH\cdot SO_3Na$, white solid, from which acetaldehyde is readily recoverable by treatment with sodium carbonate solution, (13) with hydroxylamine hydrochloride forms acetaldoxime, $CH_3\cdot CH{:}NOH$, white solid, mp 47°C, (14) with phenylhydrazine, forms acetaldehyde phenylhydrazone, $CH_3\cdot CH{:}N\cdot NH\cdot C_6H_5$, white solid, mp 98°C, (15) with magnesium methyl iodide in anhydrous ether ("Grignard's solution"), yields, after reaction with water, isopropyl alcohol, $(CH_3)_2CHOH$, a secondary alcohol, (16) with semicarbazide, forms acetaldehyde semicarbazone, $CH_3\cdot CH{:}N\cdot NH\cdot CO\cdot NH_2$, white solid, mp 162°C, (17) with chlorine, forms trichloroacetaldehyde ("chloral"), $CCl_3\cdot CHO$, (18) with H_2S, forms thioacetaldehyde, $CH_3\cdot CHS$ or $(CH_3\cdot CHS)_3$. Acetaldehyde stands chemically between ethyl alcohol on one hand—to which it can be reduced—and acetic acid on the other hand—to which it can be oxidized. These reactions of acetaldehyde, coupled with its ready formation from acetylene by mercuric sulfate solution as a catalyzer, open up a vast field of organic chemistry with acetaldehyde as raw material: acetaldehyde hydrogenated to ethyl alcohol; oxygenated to acetic acid, thence to acetone, acetic anhydride, vinyl acetate, vinyl alcohol. Acetaldehyde is also formed by the regulated oxidation of ethyl alcohol by such a reagent as sodium dichromate in H_2SO_4 (chromic sulfate also produced). Reactions (1), (3), (14), and (16) above are most commonly used in the detection of acetaldehyde. See also **Aldehydes.**

ACETAL GROUP. An organic compound of the general formula $RCH(OR')(OR'')$ is termed an *acetal* and is formed by the reaction of an aldehyde with an alcohol, usually in the presence of small amounts of acids or appropriate inorganic salts. Acetals are stable toward alkali, are volatile, insoluble in H_2O, and generally are similar structurally to ethers. Unlike ethers, acetals are hydrolyzed by acids into their respective aldehydes. $H(R)CO + (HO\cdot C_2H_5)_2 \rightarrow H(R)C(OC_2H_5)_2 + H_2O$. Representative acetals include: $CH_2(OCH_3)_2$, methylene dimethyl ether, bp 42°C; $CH_3CH(OCH_3)_2$, ethylidene dimethyl ether, bp 64°C; and $CH_3CH(OC_2H_5)_2$, ethylidene diethyl ether, bp 104°C.

ACETAL RESINS. A very useful family of thermoplastic resins, obtainable both as homopolymers and copolymers, produced mainly from formaldehyde or a formaldehyde derivative. Acetals have the highest fatigue endurance of commercial thermoplastics. In their manufacture, a variety of ionic initiators, such as tertiary amines and quaternary ammonium salts, are used to effect polymerization of formaldehyde. Chain transfer controls the molecular weight of the resulting resins.

Starting ingredients may be formaldehyde or the cyclic trimer trioxane. Both form polymers of similar properties. Boron trifluoride or other Lewis acids are used to promote polymerization where trioxane is the raw material.

Acetals provide excellent resistance to most organic compounds except when exposed for long periods at elevated temperatures. The resins have limited resistance to strong acids and oxidizing agents. The copolymers and some of the homopolymers are resistant to the action of weak bases. Normally, where resistance to burning, weathering, and radiation are required, acetals are not specified. The resins are used for cams, gears, bearings, springs, sprockets, and other mechanical parts, as well as for electrical parts, housings, and hardware.

ACETIC ACID. CH_3COOH, formula weight 60.05, colorless, acrid liquid, mp 16.7°C, bp 118.1°C, sp gr 1.049. Also known as ethanoic acid or vinegar acid, this compound is miscible with H_2O, alcohol, and ether in all proportions. Acetic acid is available commercially in several concentrations. The CH_3COOH content of glacial acetic is approximately 99.7% with H_2O the principal impurity. Reagent acetic acid generally contains 36% CH_3COOH by weight. Standard commercial aqueous solutions are 28, 56, 70, 80, 85, and 90% CH_3COOH. Acetic acid is the active ingredient in vinegar in which the content ranges from 4 to 5% CH_3COOH. Acetic acid is classified as a weak, monobasic acid. The three hydrogen atoms linked to one of the two carbon atoms are not replaceable by metals.

In addition to the large quantities of vinegar produced, acetic acid in its more concentrated forms is an important high-tonnage industrial chemical, both as a reactive raw and intermediate material for various organic syntheses and as an excellent solvent. Acetic acid is required in the production of several synthetic resins and fibers, pharmaceuticals, photographic chemicals, flavorants, and bleaching and etching compounds.

Early commercial sources of acetic acid included (1) the combined action of *Bacterium aceti* and air on ethyl alcohol in an oxidation-fermentation process: $C_2H_5OH + O_2 \rightarrow CH_3COOH + H_2O$, the same reaction which occurs when weakly alcoholic beverages, such as beer or wine, are exposed to air for a prolonged period and which turn sour because of the formation of acetic acid; and (2) the destructive distillation of wood. A number of natural vinegars still are made by fermentation and marketed as the natural product, but diluted commercially and synthetically produced acetic acid is a much more economic route to follow. The wood distillation route was phased out because of shortages of raw materials and the much more attractive economy of synthetic processes.

The most important synthetic processes are (1) the oxidation of acetaldehyde, and (2) the direct synthesis from methyl alcohol and carbon monoxide. The latter reaction must proceed under very high pressure (approximately 650 atmospheres) and at about 250°C. The reaction takes place in the liquid phase and dissolved cobaltous iodide is the catalyst. $CH_3OH + CO \rightarrow CH_3COOH$ and $CH_3OCH_3 + H_2O + 2CO \rightarrow 2CH_3COOH$. The crude acid produced first is separated from the catalyst and then dehydrated and purified in an azeotropic distillation column. The final product is approximately 99.8% pure CH_3COOH.

Acetic acid solution reacts with alkalis to form acetates, e.g., sodium acetate, calcium acetate; similarly, with some oxides, e.g., lead acetate; with carbonates, e.g., sodium acetate, calcium acetate, magnesium acetate; with some sulfides, e.g., zinc acetate, manganese acetate. Ferric acetate solution, upon boiling, yields a red precipitate of basic ferric acetate. Acetic acid solution attacks many metals, liberating hydrogen and forming acetate, e.g., magnesium, zinc, iron. Acetic acid is an important organic substance, with alcohols forming esters (acetates); with phosphorus trichloride forming acetyl chloride $CH_3{\cdot}CO{\cdot}Cl$, which is an important reagent for transfer of the acetyl (CH_3CO—) group; forming acetic anhydride, also an acetyl reagent; forming acetone and calcium carbonate when passed over lime and a catalyzer (barium carbonate) or when calcium acetate is heated; forming methane (and sodium carbonate) when sodium acetate is heated with NaOH; forming mono-, di-, tri-chloroacetic (or bromoacetic) acids by reaction with chlorine (or bromine) from which hydroxy- and amino-, aldehydic-, dibasic acids, respectively, may be made; forming acetamide when ammonium acetate is distilled. Acetic acid dissolves sulfur and phosphorus, is an important solvent for organic substances, and causes painful wounds when it comes in contact with the skin. Normal acetates are soluble, basic acetates insoluble. The latter are important in their compounds with lead, and copper ("verdigris").

A large number of acetic acid esters are important industrially, including methyl, ethyl, propyl, butyl, amyl, and cetyl acetates; glycol mono- and diacetate; glyceryl mono-, di-, and triacetate; glucose pentacetate; and cellulose tri-, tetra-, and pentacetate.

Acetates may be detected by formation of foul-smelling cacodyl (poisonous) on heating with dry arsenic trioxide. Other tests for acetate are the lanthanum nitrate test in which a blue or bluish-brown ring forms when a drop of 2.5% $La(NO)_3)_3$ solution, a drop of 0.01-*N* iodine solution, and a drop of 0.1% NH_4OH solution are added to a drop of a neutral acetate solution; the ferric chloride test, in which a reddish color is produced by the addition of 1-*N* ferric chloride solution to a neutral solution of acetate; and the ethyl acetate test, in which ethyl alcohol and H_2SO_4 are added to the acetate solution and warmed to form a colorless solution.

ACETOACETIC ESTER CONDENSATION. A class of reactions occasioned by the dehydrating power of metallic sodium or sodium ethoxide on the ethyl esters of monobasic aliphatic acids and a few other esters. It is best known in the formation of acetoacetic ester:

$$2CH_3{\cdot}COOC_2H_5 + 2CH_3{\cdot}COOC_2H_5 + 2Na \rightarrow 2CH_3{\cdot}C(ONa){:}CH{\cdot}COOC_2H_5 + 2C_2H_5OH + H_2$$

The actual course of the reaction is complex. By the action of acids the sodium may be eliminated from the first product of the reaction and the free ester obtained. This may exist in the tautomeric enol and keto forms ($CH_3{\cdot}COH{:}CH{\cdot}COOC_2H_5$ and $CH_3{\cdot}CO{\cdot}CH_2{\cdot}COOC_2H_5$).

On boiling ester with acids or alkalies it will split in two ways, the circumstances determining the nature of the main product. Thus, if moderately strong acid or weak alkali is employed, acetone is formed with very little acetic acid (ketone splitting). In the presence of strong alkalies, however, very little acetone and much acetic acid result (acid splitting). Derivatives of acetoacetic ester may be decomposed in the same fashion, and this fact is responsible for the great utility of this condensation in organic synthesis. This is also due to the reactivity of the $\cdot CH_2\cdot$ group, which reacts readily with various groups, notably halogen compounds. Usually the sodium salt of the ester is used, and the condensation is followed by decarboxylation with dilute alkali, or deacylation with concentrated alkali.

$$CH_3{\cdot}CO{\cdot}CHNa{\cdot}COOC_2H_5 + RI \rightarrow CH_3{\cdot}CO{\cdot}CHR{\cdot}COOC_2H_5 + NaI$$

$$CH_3{\cdot}CO{\cdot}CHR{\cdot}COOC_2H_5 \xrightarrow[\text{Dilute alkali}]{H_2O} CH_3{\cdot}CO{\cdot}CH_2R + C_2H_5OH + CO_2$$

$$CH_3{\cdot}CO{\cdot}CHR{\cdot}COOC_2H_5 \xrightarrow[\text{Concentrated alkali}]{2H_2O} HOOC{\cdot}CH_2{\cdot}R + C_2H_5OH + CH_3COOH$$

ACETONE. $CH_3{\cdot}CO{\cdot}CH_3$, formula weight 58.08, colorless, odorous liquid ketone, mp −94.6°C, bp 56.5°C, sp gr 0.792. Also known as dimethyl ketone or propanone, this compound is miscible in all proportions with H_2O, alcohol, or ether. Acetone is a very important solvent and is widely used in the manufacture of plastics and lacquers. For storage purposes, acetylene may be dissolved in acetone. A high-tonnage chemical, acetone is the starting ingredient or intermediate for numerous organic syntheses. Closely related, industrially important compounds are diacetone alcohol (DAA) $CH_3{\cdot}CO{\cdot}CH_2{\cdot}COH(CH_3)_2$ which is used as a solvent for cellulose acetate and nitrocellulose, as well as for various resins and gums, and as a thinner for lacquers and inking materials. Sometimes DAA is mixed with castor oil for use as a hydraulic brake fluid for which its physical properties are well suited, mp −54°C, bp 166°C, sp gr 0.938. A product known as synthetic methyl acetone is prepared by mixing acetone (50%), methyl acetate (30%), and methyl alcohol (20%) and is used widely for coagulating latex and in paint removers and lacquers.

In older industrial processes, acetone is prepared (1) by passing the vapors of acetic acid over heated lime. In a first step, calcium acetate is produced, followed by a breakdown of the acetate into acetone and calcium carbonate:

$$CH_3{\cdot}CO{\cdot}O{\cdot}Ca{\cdot}OOC{\cdot}CH_33 \rightarrow CH_3{\cdot}CO{\cdot}CH_3 + CaCO_{3;}$$

and (2) by fermentation of starches, such as maize, which produce acetone along with butyl alcohol. Modern industrial processes include (3) the use of cumene as a chargestock, in which cumene first is oxidized to cumene hydroperoxide (CHP), this followed by the decomposition of CHP into acetone and phenol; and (4) by the direct oxidation of propylene, using air and catalysts. The catalyst solution consists of copper chloride and small amounts of palladium chloride. The reaction: $CH_3CH{=}CH_2 + \frac{1}{2}O_2 \rightarrow CH_3COCH_3$. During the reaction, the palladium chloride is reduced to elemental palladium and HCl. Reoxidation is effected by cupric chloride. The cuprous chloride resulting is reoxidized during the catalyst regeneration cycle. The process is carried out under moderate pressure at about 100°C.

Acetone reacts with many chemicals in a marked manner: (1) with phosphorus pentachloride, yields acetone chloride $(CH_3)_2CCl_2$, (2) with hydrogen chloride dry, yields both mesityl oxide $CH_3COCH{:}C(CH_3)_2$, liquid, bp 132°C, and phorone $(CH_3)_2C{:}CHCOCH{:}\ C(CH_3)_2$, yellow solid, mp 28°C, (3) with concentrated H_2SO_4, yields mesitylene $C_6H_3(CH_3)_3$ (1,3,5), (4) with NH_3, yields acetone amines, e.g., diacetoneamine $C_6H_{12}ONH$, (5) with HCN, yields acetone cyanhydrin $(CH_3)_2CHOH{\cdot}CN$, readily converted into alpha-hydroxy acid $(CH_3)_2CHOH{\cdot}COOH$, (6) with sodium hydrogen sulfite, forms acetone-sodiumbisulfite $(CH_3)_2COH{\cdot}SO_3Na$, white solid, from which acetone is readily recoverable by treatment with sodium carbonate solution, (7) with hydroxylamine hydrochloride, forms acetoxime $(CH_3)_2C{:}\ NOH$, solid, mp 60°C, (8) with phenylhydrazine, yields acetonephenylhydrazone $(CH_3)_2C{:}NNHC_6H_5{\cdot}H_2O$, solid, mp 16°C, anhydrous compound, mp 42°C, (9) with semicarbazide, forms acetonesemicarbazone $(CH_3)C{:}\ NNHCONH_2$, solid, mp 189°C, (10) with magnesium methyl iodide in anhydrous ether ("Grignard's solution"), yields, after reaction with H_2O, trimethylcarbinol $(CH_3)_3COH$, a tertiary alcohol, (11) with ethyl thioalcohol and hydrogen chloride dry, yields mercaptol $(CH_3)_2C(SC_2H_5)_2$, (12) with hypochlorite, hypobromite, or hypoiodite solution, yields chloroform $CHCl_3$, bromoform $CHBr_3$ or iodoform CHI_3, respectively, (13) with most reducing agents, forms isopropyl alcohol $(CH_3)_2CHOH$, a secondary alcohol, but with sodium amalgam forms pinacone $(CH_3)_2COH{\cdot}COH(CH_3)_2$, (14) with sodium dichromate and H_2SO_4, forms acetic acid CH_3COOH plus CO_2. When acetone vapor is passed through a tube at a dull red heat, ketene $CH_2{:}CO$ and methane CH_4 are formed.

ACETYLENE. $CH{:}CH$, formula weight 26.04, mp −81.5°C, bp −84°C, sp gr 0.905 (air = 1.000). Sometimes referred to as *ethyne*, *ethine*, or *gaseous carbon* (92.3% of the compound is C), acetylene is moderately soluble in H_2O or alcohol, and exceptionally soluble in acetone (300 volumes of acetylene in 1 volume of acetone at 12 atmospheres pressure). The gas burns when ignited in air with a luminous sooty flame, requiring a specially devised burner for illumination purposes. An explosive mixture is formed with air over a wide range (about 3 to 80% acetylene), but safe handling is improved when the gas is dissolved in acetone. The heating value is 1455 Btu/ft^3 (8.9 Cal/m^3).

Although acetylene still is used in a number of organic syntheses on an industrial scale, its use on a high-tonnage basis has diminished because of the lower cost of other starting materials, such as ethylene and propylene. Acetylene has been widely used in the production of halogen derivatives, acrylonitrile, acetaldehyde, and vinyl chloride. Within recent years, producers of acrylonitrile switched to propylene as a starting material.

Commercially, acetylene is produced from the pyrolysis of naphtha in a two-stage cracking process. Both acetylene and ethylene are end-products. The ratio of the two products can be changed by varying the naphtha feed rate. Acetylene also has been produced by a submerged-flame process from crude oil. In essence, gasification of the crude oil occurs by means of the flame, which is supported by oxygen beneath the surface of the oil. Combustion and cracking of the oil take place at the boundaries of the flame. The composition of the cracked gas includes about 6.3% acetylene and 6.7% ethylene. Thus, further separation and purification are required. Several years ago when procedures were developed for the safe handling of acetylene on a large scale, J. W. Reppe worked out a series of reactions that later became known as "Reppe chemistry." These reactions were particularly important to the manufacture of many high polymers and other synthetic products. Reppe and his associates were able to effect synthesis of chemicals that had been commercially unavailable. An example is the synthesis of cyclooctatetraene by heating a solution of acetylene under pressure in tetrahydrofuran in the presence of a nickel cyanide catalyst. In another reaction, acrylic acid was produced from CO and H_2O in the presence of a nickel catalyst: $C_2H_2 + CO + H_2O \rightarrow CH_2{:}CH{\cdot}COOH$. These two reactions are representative of a much larger number of reactions, both those that are straight-chain only, and those involving ring closure.

Acetylene reacts (1) with chlorine, to form acetylene tetrachloride $C_2H_2Cl_4$ or $CHCl_2{\cdot}CHCl_2$ or acetylene dichloride $C_2H_2Cl_2$ or $CHCl{:}CHCl$, (2) with bromine, to form acetylene tetrabromide $C_2H_2Br_4$ or $CHBr_2{\cdot}CHBr_2$ or acetylene dibromide $C_2H_2Br_2$ or $CHBr{:}CHBr$, (3) with hydrogen chloride (bromide, iodide), to form ethylene monochloride $CH_2{:}CHCl$ (monobromide, monoiodide), and 1,1-dichloroethane, ethylidene chloride $CH_3{\cdot}CHCl_2$ (dibromide, diiodide), (4) with H_2O in the presence of a catalyzer, e.g., mercuric sulfate, to form acetaldehyde $CH_3{\cdot}CHO$, (5) with hydrogen, in the presence of a catalyzer, e.g., finely divided nickel heated, to form ethylene C_2H_4 or ethane C_2H_6, (6) with metals, such as copper or nickel, when moist, also lead or zinc, when moist and unpurified. Tin is not attacked. Sodium yields, upon heating, the compounds C_2HNa and C_2Na_2. (7) With ammoniocuprous (or silver) salt solution, to form cuprous (or silver) acetylide C_2Cu_2, dark red precipitate, explosive when dry, and yielding acetylene upon treatment with acid, (8) with mercuric chloride solution, to form trichloromercuric acetaldehyde $C(HgCl)_3{\cdot}CHO$, precipitate, which yields with HCl acetaldehyde plus mercuric chloride.

ACETYLENE SERIES. A series of unsaturated hydrocarbons having the general formula C_nH_{2n-2}, and containing a triple bond between two carbon atoms. The series is named after the simplest compound of the series, acetylene $HC{:}CH$. In more modern terminology, this series of compounds is termed the *alkynes*. See also **Alkynes.**

ACHENE. A single-seeded, indehiscent fruit, in which the seed is free from the ovary wall except at the point of attachment. An example is the sunflower "seed."

ACHERNAR (α Eridani). Ranking tenth in apparent brightness among the stars, Achernar has a true brightness value of 200 as compared with unity for the sun. Achernar is a blue-white, spectral B type star and is one of the end stars in the constellation Eridanus, located south of the ecliptic and in the viewing vicinity of the Magellanic Clouds. Estimated distance from the earth is 65 light years. See also **Constellations;** and **Star.**

ACHILLES TENDON. In humans, the prominent tendon at the back of the ankle, extending from the muscle of the calf to the heel. Technically, it is the tendon which attaches the gastrocnemius and soleus muscles to the calcaneus or heel bone. The name derived from human anat-

Achilles tendon.

omy is used in relation to other vertebrates. See accompanying illustration.

ACHLORHYDRIA. The cessation of acid production by the stomach. The low pH of the normal stomach contents is a barrier to infection by various organisms and, where achlorhydria develops—particularly in malnutrition—it renders the patient more susceptible to infection, such as by *Vibrio cholerae* and *Giardia lamblia*. The condition is relatively common among people of about 50 years of age and older, affecting 15 to 20% of the population in this age group. The acid deficiency also occurs in about 30% of patients with adult onset-type of primary hypogamma-globulinemia. A well-balanced diet of easily digestible foods minimizes the discomforting effects of complete absence of hydrochloric acid in the stomach. The condition does not preclude full digestion of fats and proteins, the latter being attacked by intestinal and pancreatic enzymes. In rare cases, where diarrhea may result from achlorhydria, dilute hydrochloric acid may be administered by mouth. Where this causes an increase in discomfort or even pain, the use of dexamethasone or mucosal coating agents is preferred.

Commonly, achlorhydria may not be accompanied by other diseases, but in some cases there is a connection. For example, achlorhydria is an abnormality which sometimes occurs with severe iron deficiency. Histalog-fast achlorhydria, resulting from intrinsic factor deficiency in gastric juice, may be an indication of pernicious anemia. Hyperplastic polyps are often found in association with achlorhydria.

Excessive alcohol intake can also lead to achlorhydria and it is said that the resistance for cyanide poisoning of the Russian mystic, Rasputin, was attributable to that effect. The great amount of vodka which he consumed led to achlorhydria and thus the ingested potassium cyanide did not liberate lethal hydrocyanic gas, nor was the potassium salt absorbed through the stomach walls.

R. C. V.

ACHOLIA. Absence or lack of secretion of bile.

ACHONDRITES. A form of stony meteorites without chondri, and having textures similar to those of some terrestrial rocks. See **Meteoroids and Meteorites.**

ACHROMAT. A compound lens corrected so as to have the same focal length for two or more different wavelengths. Commonly, the F- and C-lines are the chosen wavelengths. See **Fraunhofer Lines.**

Achromats are used in optical microscopes for routine work, although they show some field curvature.

ACHROMATIC. Free from hue. Transmitting light without showing its constituent colors, or separating it into them. An achromatic color sometimes is referred to as gray.

Achromatic Combination. If reversed crown and flint prisms are made of such angle that the angles of dispersion between any two different wavelengths of light are alike but reversed in direction, then these two colors will not be separated and all colors lying between them will be separated little, if any, from each other. By using three kinds of glass, it is possible to bring three colors together. When the dispersions balance, the angles of deviation will, in general, not balance. This same principle is used in making achromatic lenses. Achromatic prisms have a maximum of deviation and a minimum of dispersion, whereas an Amici prism disperses the light with a minimum of deviation.

Achromatic Locus. Chromaticities that may be acceptable reference standards under circumstances of common occurrence are represented in a chromaticity diagram by points in a region which may be called the "achromatic locus." Any point within the achromatic locus, chosen as a reference point, may be called an "achromatic point." Such points have also been called "white points." However, the term "white point" is best used to specify the intersection of the various achromatic loci obtained under different conditions of adaptation.

Achromatic Stimulus. A visual stimulus that is capable of exciting a color sensation of no hue. In practice, an arbitrarily chosen chromaticity, such as that of the prevailing illumination.

ACID-BASE REGULATION (Blood). The hydrogen ion concentration of the blood is maintained at a constant level of pH 7.4 by a complex system of physico-chemical processes, involving, among others, neutralization, buffering, and excretion by the lungs and kidneys. This topic is sometimes referred to as *acid-base metabolism*. The clinical importance of acids and bases in life processes derives from several fundamental factors. (1) Most chemical reactions within the body take place in water solutions. The type and rate of such reactions is seriously affected by acid-base concentrations, of which pH is one indication. (2) Hydrogen ions are mobile charged particles and the distribution of such ions as sodium, potassium, and chloride in the cell environment are ultimately affected by hydrogen ion concentration (pH). (3) It also has been established that hydrogen ion concentration influences the three-dimensional configurations of proteins. Protein conformational changes affect the biochemical activity of proteins and thus can affect normal protein function. For example, enzymes, a particular class of proteins, exhibit optimal activity within a narrow range of pH. Most physiological activities, and especially muscular exercise, are accompanied by the production of acid, to neutralize which, a substantial alkali reserve, mainly in the form of bicarbonate, is maintained in the plasma, and so long as the ratio of carbon dioxide to bicarbonate remains constant, the hydrogen ion concentration of the blood does not alter. Any non-volatile acid, such as lactic or phosphoric, entering the blood reacts with the bicarbonate of the alkali reserve to form carbon dioxide, which is volatile, and which combines with hemoglobin by which it is transported to the lungs and eliminated by the processes of respiration. It will also be evident from this that no acid stronger than carbon dioxide can exist in the blood. The foregoing neutralizing and buffering effects of bicarbonate and hemoglobin are short-term effects; to insure final elimination of excess acid or alkali, certain vital reactions come into play. The rate and depth of respiration are governed by the level of carbon dioxide in the blood, through the action of the respiratory center in the brain; by this means the pulmonary ventilation rate is continually adjusted to secure adequate elimination of carbon dioxide. In the kidneys two mechanisms operate; ammonia is formed, whereby acidic substances in process of excretion are neutralized, setting free basic ions such as sodium to return to the blood to help maintain the alkali reserve. Where there is a tendency to the development of increased acidity in the blood, the kidneys are able selectively to re-absorb sodium bicarbonate from the urine being excreted, and to release into it acid sodium phosphate; where there is a tendency to alkalemia, alkaline sodium phosphate is excreted, the hydrogen ions thus liberated being re-absorbed to restore the diminishing hydrogen ion concentration. See also **Achlorhydria; Acidosis; Alkalosis; Blood; pH (Hydrogen ion concentration); Potassium** and **Sodium (In Biological Systems).**

ACIDIC SOLVENT. A solvent which is strongly protogenic, i.e., which has a strong tendency to donate protons and little tendency to accept them. Liquid hydrogen chloride and hydrogen fluoride are acidic solvents, and in them even such normally strong acids as nitric acid do not exhibit acidic properties, since there are no molecules which can accept protons; but, on the contrary, behave to some extent as bases by accepting protons yielded by the dissociation of the HCl or the HF. See **Acids and Bases.**

ACIDIMETRY. An analytical method for determining the quantity of acid in a given sample by titration against a standard solution of a base, or, more broadly, a method of analysis by titration where the end point is recognized by a change in pH (hydrogen ion concentration). See also **Analysis (Chemical); pH (Hydrogen Ion Concentration; Titration (Potentiometric);** and **Titration (Thermometric).**

ACIDITY. The amount of acid present, expressed for a solution either as the molecular concentration of acid, in terms of normality, molality, etc., or the ionic concentration (hydrogen ions or protons) in terms of pH (the logarithm of the reciprocal of the hydrogen ion concentration). The acidity of a base is the number of molecules of monoatomic acid which one molecule of the base can neutralize. See **Acids and Bases.**

ACID NUMBER. A term used in the analysis of fats or waxes to designate the number of milligrams of potassium hydroxide required to neutralize the free fatty acids in 1 gram of substance. The determination is performed by titrating an alcoholic solution of the wax or fat with tenth or half-normal alkali, using phenolphthalein as indicator.

ACIDOSIS. A condition of excess acidity (or depletion of alkali) in the body, in which acids are absorbed or formed in excess of their elimination, thus increasing the hydrogen ion concentration of the blood, exceeding the normal limit of 7.4. The acidity-alkalinity ratio in body tissue normally is delicately controlled by several mechanisms, notably the regulation of carbon dioxide-oxygen transfer in the lungs, the presence of buffer compounds in the blood, and the numerous sensing areas that are a part of the central nervous system. Normally, acidic materials are produced in excess in the body, this excess being neutralized by the presence of free alkaline elements, such as sodium occurring in plasma. The combination of sodium with excess acids produces carbon dioxide which is exhaled. Acidosis may result from: (1) severe exercise, leading to increased carbon dioxide content of the blood, (2) sleep, especially under narcosis, where the elimination of carbon dioxide is depressed, (3) heart failure, where there is diminished ventilation of carbon dioxide through the lungs, (4) diabetes and starvation, in which organic acids, such as β-hydroxybutyric and acetoacetic acids, accumulate, (5) kidney failure, in which the damaged kidneys cannot excrete acid radicals, and (6) severe diarrhea, in which there is loss of alkaline substances. Nausea, vomiting, and weakness sometimes may accompany acidosis. See also **Acid-Base Regulation (Blood); Blood; Kidney and Urinary Tract;** and **Potassium and Sodium (In Biological Systems).**

ACID RAIN. See **Climatology; Electric Power Production and Distribution; Pollution (Air).**

ACIDS AND BASES. The conventional definition of an acid is that it is an electrolyte that furnishes protons, i.e., hydrogen ions, H^+. An acid is sour to the taste and usually quite corrosive. A base is an electrolyte that furnishes hydroxyl ions, OH^-. A base is bitter to the taste and also usually quite corrosive. These definitions were formulated in terms of water solutions and, consequently, do not embrace situations where some ionizing medium other than water may be involved. In the definition of Lowry and Brønsted, an acid is a proton donor and a base is a proton acceptor. Acid-base theory is described later.

Acidification is the operation of creating an excess of hydrogen ions, normally involving the addition of an acid to a neutral or alkaline solution until a pH below 7 is achieved, thus indicating an excess of hydrogen ions. In *neutralization*, a balance between hydrogen and hydroxyl ions is effected. An acid solution may be neutralized by the addition of a base; and vice versa. The products of neutralization are a salt and water.

Some of the inorganic acids, such as hydrochloric acid, HCl, nitric acid, HNO_3, and sulfuric acid, H_2SO_4, are very-high-tonnage products and are considered very important chemical raw materials. The most common inorganic bases (or alkalis) include sodium hydroxide, NaOH, and potassium hydroxide, KOH, and also are high-tonnage materials, particularly NaOH.

Several classes of organic substances are classified as acids, notably the carboxylic acids, the amino acids, and the nucleic acids. These and the previously mentioned materials are described elsewhere in this volume.

Principal theories of acids and bases have included: (1) Arrhenius-Ostwald theory, which was proposed soon after the concept of the *ionization* of chemical substances in aqueous solutions was generally accepted. (2) Much later (1923), J. N. Brønsted defined an acid as a source of protons and a base is an acceptor of protons. (3) T. M. Lowry, working in the same time frame as Brønsted, developed a similar concept and, over the years, the concept has been referred to in the literature as the *Lowry-Brønsted theory*. It will be noted that this theory altered the definition of an acid very little, continuing to emphasize the role of the hydrogen ion. However, the definition of a base was extended beyond the role of the hydroxyl ion to include a wide variety of uncharged species, such as ammonia and the amines. (4) In 1938, G. N. Lewis further broadened the definition of Lowry-Brønsted. Lewis defined an acid as anything which can attach itself to something with an unshared pair of electrons. The broad definition of Lewis creates some difficulties when one attempts to categorize Lewis acids and bases. R. G. Pearson (1963) suggested two main categories—hard and soft acids as well as hard and soft bases. These are described in more detail by Long and Boyd (1983). (5) In 1939, M. Usanovich proposed still another theory called the *positive-negative theory*, also developed in detail by Long and Boyd.

In terms of the definition that an acid is a proton donor and a base is a proton acceptor, hydrochloric acid, water, and ammonia (NH_3) are acids in the reactions:

$$HCl \rightleftarrows H^+ + Cl^-$$
$$H_2O \rightleftarrows H^+ + OH^-$$
$$NH_3 \rightleftarrows H^+ + NH_2^-$$

Note that this definition is different in at least two major respects from the conventional definition of an acid as a substance dissociating to give H^+ in water. The Lowry-Brønsted definition states that for every acid there be a "conjugate" base, and vice versa. Thus, in the examples cited above, Cl^-, OH^-, and NH^- are the conjugate bases of HCl, H_2O, and NH_3. Furthermore, since the equations given above should more properly be written:

$$HCl + H_2O \rightleftarrows H_3O^+ + Cl^-$$
$$H_2O + H_2O \rightleftarrows H_3O^+ + OH^-$$
$$NH_3 + H_2O \rightleftarrows H_3O^+ + NH_2^-$$

It can be seen that every acid-base reaction involving transfer of a proton will involve two conjugate acid-base pairs, e.g., in the last equation NH_3 and H_3O^+ are the acids and NH_2^- and H_2O the respective conjugate bases. On the other hand, in the reaction:

$$NH_3 + H_2O \rightleftarrows NH_4^+ + OH^-$$

H_2O and NH_4^- are the acids and NH_3 and OH^- the bases. In other reactions, e.g.,

Base$_1$	*Acid*$_2$	*Acid*$_1$	*Base*$_2$
$C_2H_3O_2^-$	$+ H_2O$	$\rightleftarrows HC_2H_3O_2$	$+ OH^-$
HCO_3^-	$+ HCO_3^-$	$\rightleftarrows H_2CO_3$	$+ CO_3^{-2}$
$N_2H_5^+$	$+ N_2H_5^+$	$\rightleftarrows N_2H_6^{+2}$	$+ N_2H_4$
H_2O	$+ Cr(H_2O)_6^{+3}$	$\rightleftarrows N_3O^+$	$+ Cr(H_2O)_5\ OH^{2+}$

the conjugate acids and bases are as indicated. The theory is not limited to the aqueous solution; for example, the following reactions can be considered in exactly the same light:

Base$_1$	*Acid*$_2$	*Acid*$_1$	*Base*$_2$
NH_3	$+ HCl$	$\rightleftarrows NH_4^+$	$+ Cl^-$
CH_3CO_2H	$+ HF$	$\rightleftarrows CH_3CO_2H_2^+$	$+ F^-$
HF	$+ HClO_4$	$\rightleftarrows H_2F^+$	$+ ClO_4^-$
$(CH_3)_2O$	$+ HI$	$\rightleftarrows (CH_3)_2\ OH^+$	$+ I^-$
C_6H_6	$+ HSO_3F$	$\rightleftarrows C_6H_7^+$	$+ SO_3F^-$

Acids may be classified according to their charge or lack of it. Thus, in the reactions cited above, there are "molecular" acids and bases, such as HCl, H_2CO_3, $HClO_4$, etc., and N_2H_4, $(CH_3)_2O$, C_6H_6, etc., and also cationic acids and bases, such as H_3O^+, $N_2H_5^+$, $N_2H_6^{2+}$, NH_4^+, $(CH_3)_2OH^+$, etc., as well as anionic acids and bases, such as HCO_3^-, Cl^-, NH_2^-, NH_3^{-2}, etc. In a more general definition, Lewis calls a base any substance with a free pair of electrons that it is capable of sharing with an electron pair acceptor, which is called an acid. For example, in the reaction:

$$(C_2H_5)_2O\!: + BF_3 \rightarrow (C_2H_5)_2O\!:\!BF_3$$

the ethyl ether molecule is called a base, the boron trifluroide, an acid. The complex is called a *Lewis salt*, or *addition compound*.

Acids are classified as monobasic, dibasic, tribasic, polybasic, etc., according to the number (one, two, three, several, etc.) of hydrogen atoms, replaceable by bases, contained in a molecule. They are further classified as (1) organic when the molecule contains carbon; (1a) carboxylic, when the proton is from a —COOH group; (2) normal, if they are derived from phosphorus or arsenic, and contain three hydroxyl groups: (3) ortho, meta, or para, according to the location of the carboxyl group in relation to another substituent in a cyclic compound; or (4) ortho, meta, or pyro, according to their composition.

Superacids. Although mentioned in the literature as early as 1927, *superacids* were not investigated aggressively until the 1970s. Prior to the concept of superacids, scientists generally regarded the familiar mineral acids (HF, HNO_3, H_2SO_4, etc.) as the strongest acids attainable. Relatively recently, acidities up to 10^{12} times that of H_2SO_4 have been produced.

In very highly concentrated acid solutions, the commonly used measurement of pH is not applicable. See also **pH (Hydrogen Ion Concentration)**. Rather, the acidity must be related to the degree of transformation of a base with its conjugate acid. In the *Hammett acidity function*, developed by Hammett and Deyrup in 1932,

$$H_0 = pK_{BH^+} - \log \frac{BH^+}{B}$$

where $pkBH^+$ is the dissociation constant of the conjugate acid (BH^+), and BH^+/B is the ionization ratio, measurable by spectroscopic means (UV or NMR). In the Hammett acidity function, acidity is a logarithmic scale wherein H_2SO_4 (100%) has an H_0 of -11.9; and HF, an H_0 of -11.0.

As pointed out by Olah et al. (1979), "The acidity of a sulfuric acid solution can be increased by the addition of solutes that behave as acids in the system: $HA + H_2SO_4 \rightleftarrows H_3SO_4^+ + A^-$. These solutes increase the concentration of the highly acidic H_3SO_4 cation just as the addition of an acid to water increases the concentration of the oxonium ion. H_3O^+. Fuming sulfuric acid (oleum) contains a series of such acids, the polysulfuric acids, the simplest of which is disulfuric acid, $H_2S_2O_7$, which ionizes as a moderately strong acid in sulfuric acid: $H_2S_2O_7 + H_2SO_4 \rightleftarrows H_3SO_4^- + HS_2O_7^-$. Higher polysulfuric acids, such as $H_2S_3O_{10}$ and $H_2S_4O_{13}$, also behave as acids and appear somewhat stronger than $H_2S_2O_7$."

Hull and Conant in 1927 showed that weak organic bases (ketones and aldehydes) will form salts with perchloric acid in nonaqueous solvents. This results from the ability of perchloric acid in nonaqueous systems to protonate these weak bases. These early investigators called such a system a superacid. Some authorities believe that any protic acid that is stronger than sulfuric acid (100%) should be typed as a superacid. Based upon this criterion, fluorosulfuric acid and trifluoromethanesulfonic acid, among others, are so classified. Acidic oxides (silica and silica-alumina) have been used as solid acid catalysts for many years. Within the last few years, solid acid systems of considerably greater strength have been developed and can be classified as *solid superacids*.

Superacids have found a number of practical uses. Fluoroantimonic acid, sometimes called *Magic Acid*, is particularly effective in preparing stable, long-lived carbocations. Such substances are too reactive to exist as stable species in less acidic solvents. These acids permit the protonation of very weak bases. For example, superacids, such as Magic Acid, can protonate saturated hydrocarbons (alkanes) and thus can play an important role in the chemical transformation of hydrocarbons, including the processes of isomerization and alkylation. See also **Alkylation;** and **Isomerization.** Superacids also can play key roles in polymerization and in various organic syntheses involving dienone-phenol rearrangement, reduction, carbonylation, oxidation, among others. Superacids also play a role in inorganic chemistry, notably in the case of halogen cations and the cations of nonmetallic elements, such as sulfur, selenium, and tellurium.

Free Hydroxyl Radical. It is important to distinguish the free radical ·OH and the OH^- ion previously mentioned. The free radical is created by complex reactions of so-called "excited" oxygen with hydrogen as the result of exposure to solar ultraviolet light. The radical has been found to be an important factor in atmospheric and oceanic chemistry. The life span of the radical is but a second or two, during which time it reacts with numerous atmospheric pollutants in a scavenging (oxidizing) manner. For example, it reacts with carbon monoxide, as commonly encountered in atmospheric smog. It also reacts with sulfurous gases and with hydrocarbons, as may result from incomplete combustion processes or that have escaped into the atmosphere (because of their volatility) from various sources.

Because of the heavy workload placed upon the hydroxyl radical through such "cleansing" reactions in the atmosphere, some scientists are concerned that the atmospheric content of ·OH has diminished with increasing pollution, estimating the probable drop to be as much as 5–25% during the past three centuries since the start of the Industrial Revolution. Ironically, some of the very pollutants that are targets for reduction also are compounds from which the ·OH radical is produced and, as they are reduced, so will the concentration of ·OH be reduced. The fact that there is only one hydroxyl radical per trillion air molecules must not detract from its effectiveness as a scavenger.

Scientists at the Georgia Institute of Technology have devised a mass spectrometric means for testing the various theories pertaining to the chemistry of ·OH.

The probable importance of ·OH in the oceans also is being investigated. Researchers at Washington State University and the Brookhaven National Laboratory have confirmed the presence of ·OH in seawater and now are attempting to measure its content quantitatively and to determine the sources of its formation. Dissolved organic matter is one highly suspected source. Tentatively, it has been concluded (using a method called flash photolysis) that ·OH concentrations (as well as daughter radicals) range from 5 to 15 times higher in deep water than in open-ocean surface waters. This may indicate that ·OH may have some impact on biota residing in deep water and may enhance the secondary production of bacterial growth, particularly in "carbon limited" oligotrophic waters, in upwelling waters, and in regions with high ultraviolet radiation.

See also specific acids and bases, such as sulfuric acid and sodium hydroxide, in alphabetically arranged entries throughout this *Encyclopedia.*

Additional Reading

Kerr, R. A.: "Hydroxyl, the Cleanser That Thrives on Dirt," *Science*, 1210 (September 13, 1991).

Mopper, K., and X. Zhou: "Hydroxyl Radical Photoproduction in the Sea and Its Potential Impact on Marine Processes," *Science*, 661 (November 2, 1990).

Mortimer, C. E.: "Chemistry—A Conceptual Approach," 4th Edition, Van Nostrand Reinhold, New York, 1979.

Staff: "Handbook of Chemistry and Physics," 73rd Edition, CRC Press, Boca Raton, Florida, 1992–1993.

Walling, C.: "Fifty Years of Free Radicals," American Chemical Society, Washington, D.C., 1992.

ACIDULANTS AND ALKALIZERS (Foods). Well over 50 chemical additives are commonly used in food processing or as ingredients of final food products, essentially to control the pH (hydrogen ion concentration) of the process and/or product. An excess of hydrogen ions, as contributed by acid substances, produces a sour taste, whereas an excess of hydroxyl ions, as contributed by alkaline substances, creates a bitter taste. Soft drinks and instant fruit drinks, for example, owe their tart flavor to acidic substances, such as citric acid. Certain candies, chewing gums, jellies, jams, and salad dressings are among the many other products where a certain degree of tartness contributes to the overall taste and appeal.

Taste is only one of several qualities of a process or product which is affected by an excess of either of these ions. Some raw materials are naturally too acidic, others too alkaline—so that neutralizers must be added to adjust the pH within an acceptable range. In the dairy industry, for example, the acid in sour cream must be adjusted by the addition of alkaline compounds in order that satisfactory butter can be churned. Quite often, the pH may be difficult to adjust or to maintain after adjustment. Stability of pH can be accomplished by the addition of buffering agents which, within limits, effectively maintain the desired pH even when additional acid or alkali is added. For example, orange-flavored instant breakfast drink has just enough "bite" from the addition of potassium citrate (a buffering agent) to regulate the tart flavor imparted by another ingredient, citric acid. In some instances, the presence of acids or alkalies assist mechanical processing operations in food preparation. Acids, for example, make it easier to peel fruits and tubers. Alkaline solutions are widely used in dehairing animal carcasses.

The pH values of various food substances cover a wide range. Plant tissues and fluids (about 5.2); animal tissues and fluids (about 7.0 to 7.5); lemon juice (2.0 to 2.2); acid fruits (3.0 to 4.5); fruit jellies (3.0 to 3.5).

Acidulants commonly used in food processing include: Acetic acid (glacial), citric acid, fumaric acid, glucono delta-lactone, hydrochloric acid, lactic acid, malic acid, phosphoric acid, potassium acid tartrate, sulfuric acid, and tartaric acid. Alkalies commonly used include: Ammonium bicarbonate, ammonium hydroxide, calcium carbonate, calcium oxide, magnesium carbonate, magnesium hydroxide, magnesium oxide, potassium bicarbonate, potassium carbonate, potassium hydroxide, sodium bicarbonate, sodium carbonate, sodium hydroxide, and sodium sesquicarbonate. Among the buffers and neutralizing agents favored are: Adipic acid, aluminum ammonium sulfate, ammonium phosphate (di- or monobasic), calcium citrate, calcium gluconate, sodium acid pyrophosphate, sodium phosphate (di-, mono-, and tri-basic), sodium pyrophosphate, and succinic acid.

See also **Buffer (Chemical)**; and **pH (Hydrogen Ion Concentration).**

Functions of Acidulants. In the *baking industry*, acidulants and their salts control pH to inhibit spoilage by microbial actions to enhance the stability of foams (such as whipped egg albumin), to assist in leavening in order to achieve desired volume and flavoring, and to maximize the performance of artificial preservatives. A variety of the food acids previously mentioned is used. For example, citric acid traditionally has been favored by bakers for pie fillings. Baking powders (leavening agents) frequently will contain adipic acid, fumaric acid, and cream of tartar. Fumaric acid, in particular, has been the choice for leavening systems of cakes, pancakes, biscuits, waffles, crackers, cookies, and doughnuts. This acid also provides the desired characteristic flavor for sour rye bread—this eliminating fermentation of the dough to achieve desired flavor. Lactic acid and its salts sometimes are used as dough conditioners.

Acidulants are used in the *soft drink beverage industry* for producing a tart taste, improving flavor balance, modifying the "sweetness" provided by sugar and other sweeteners, extending shelf life by reducing pH value of final product, and improving the performance of antimicrobial agents. Specific acidulants preferred vary with the type of beverage—i.e. carbonated, noncarbonated, dry (reconstituted by addition of water), and low-calorie products.

In the production of *confections and gelatin desserts*, acidulants are used mainly for enhancing flavor, maintaining viscosity, and controlling gel formation. In confections, such as hard candies, acidulants are used to increase tartness and to enhance fruit flavors. Acidulants also contribute to the ease of manufacturing.

In *dairy products*, acidulants, in addition to achieving many of the foregoing functions, also help to process the products. As an example, adipic acid improves the texture and melting characteristics of processed cheese and cheese foods, where pH control is very important.

In *fruit and vegetable processing*, acidulants play somewhat different roles than previously described. These would include reducing process heating requirements through pH control, inactivating certain enzymes that reduce shelf life, and chelation of trace metals that may be present (through catalytic enzymatic oxidation). Citric acid is used widely in canned fruits, such as apricots, peaches, pears, cherries, applesauce, and grapes, to retain the firmness of the products during processing. The acid also provides a desirable tartness in the final products.

In the *processed meat field*, citric acid, along with oxidants, is used to prevent rancidity in frankfurters and sausages. Sodium citrate is used in processing livestock blood, which is used to manufacture some sausages and pet foods.

Acidulants and alkalizers, like other food additives, are controlled by regulatory bodies in most industrial nations. Some of the additives mentioned in this article are considered to be "Generally Regarded as Safe," having a GRAS classification. These include acetic, adipic, citric, glucone delta lactone, lactic, malic, phosphoric, and tartaric acids. Others are covered by the Code of Federal Registration (FDA) in the United States.

A very orderly and informative article (Dziezak 1990) is suggested as a source of detailed information on this topic.

Additional Reading

Bouchard, E. F., and E. G. Merritt: "Citric Acid" in "Encyclopedia of Chemical Technology," Kirk-Othmer, Editors, Vol. 6, Wiley, New York, 1979.

Dziezak, J. D.: "Acidulants: Ingredients That Do More than Meet the Acid Test," *Food Techy.*, 76 (January 1990).

Fellows, P. J.: "Food Processing Technology," VCH Publishers, New York, 1988.

Igoe, R. S.: "Dictionary of Food Ingredients," Van Nostrand Reinhold, New York, 1989.

Lewis, M. J.: "Physical Properties of Foods and Food Processing Systems," VCH Publishers, New York, 1987.

Ockerman, H. W.: "Sausage and Processed Meat Formulations," Van Nostrand Reinhold, New York, 1989.

Porter, A.: "Spice Up Your Candy with Citric Acid," *Candy Industry*, **150**(08): 34 (1985).

Vreeman, G.: "Lactic Acid Accents Continuous Candy Lines," *Candy Industry* **15**(1): 33 (1986).

ACLINIC LINE (or Dip Equator; Magnetic Equator). The line through those points on the earth's surface at which the magnetic inclination is zero. The aclinic line is a particular case of an isoclinic line.

ACMITE-AEGERINE. Acmite is a comparatively rare rockmaking mineral, usually found in nephelite syenites or other nephelite or leucite-bearing rocks, as phonolites. Chemically, it is a soda-iron silicate, and its name refers to its sharply pointed monoclinic crystals. Bluntly terminated crystals form the variety aegerine, named for Aegir, the Icelandic sea god.

Acmite has a hardness of 6 to 6.5, specific gravity 3.5, vitreous; color brown to greenish-black (aegerine), or red-brown to dark green and black (acmite). Acmite is synonymous with aegerine, but usually restricted to the long slender crystalline variety of brown color.

The original acmite locality is in Greenland. Norway, the former U.S.S.R., Kenya, India, and Mt. St. Hilaire, Quebec, Canada furnish fine specimens. United States localities are Magnet Cove, Arkansas, and Libby, Montana, where a variety carrying vanadium occurs.

ACNE VULGARIS. A chronic disorder of the pilosebaceous units and generally confined to the face, chest, and back. The primary lesions are horny plugs (blackheads), which later develop into pink papules, pustules, or nodules. Nodules are tender, acute, and localized collections of pus deep in the dermis. Large pustular lesions may develop and may break down adjacent tissue to form lakes of pus, sinuses, and characteristically pitted scars. Studies have shown that heredity is a major predisposing factor in the disease. It has been described by some authorities as polygenic (summed effects of many genes) and thus difficult to clearly delineate in terms of heredity.

Prior to puberty, the small pilosebaceous units are dormant. At puberty, these units enlarge and produce sebum by action of the sebaceous gland which converts circulating testosterone to 5α-dihydrostestosterone, which is the tissue androgen. See **Androgens.** Acne may occur when an excess of sebum is produced; or when the pilosebaceous follicular openings are too small to allow increased sebum flow; or when both conditions are present. The duration of acne vulgaris is highly variable and can persist well into the third decade of life. But, normally the disease peaks during the teen years and early twenties.

The microflora of acne lesions are usually those that are found in normal skin, including the anaerobic diphtheroid *Propionibacterium acne* and coagulase-negative staphylococci and micrococci. The yeast *Pityrosporum* usually colonizes the superficial layers of the lesions. Acneiform eruptions also may be caused by certain drugs, including phenytoin, isoniazid, lithium, and drugs containing halogens. Symptoms of steroid acne usually occur within two weeks after the commencement of systemic steroids. Papulopustular eruptions congregate on the upper chest, back, and arms. *Chloracne* is the result of occupational exposure to chlorinated hydrocarbons, tars, and cutting oils. Eruptions usually occur at points of contact with these foreign materials. *Acne rosacea* tends to affect middle-aged adults and occurs on the forehead, cheeks, nose, and chin, usually with the presence of facial flushing.

General measures to be taken for the care of acne include: (1) Thorough cleansing of the skin with soap and water at least once per day to remove dirt, sebum, and scales. Greasy cosmetics are to be avoided, but a water-base makeup may be used. (2) Frequent cleansing of scalp and hair with an antiseborrheic shampoo for removal of excess oil. (3) Hair should be worn off the face. (4) Physical tampering, such as squeezing of the lesions should be avoided. (5) Mild exposure to the sun's rays may be beneficial, but should be avoided in persons using topical retinoic acid. (6) The patient should attempt to relate types of food intake with any exacerbation of the acne—because some authorities believe there are dietary factors involved. However, formerly held associations of acne with certain taboos, such as chocolate, coffee, cola drinks, fatty foods, nuts, ice cream, and sweets, have not been proven. Some lotions and creams have been helpful in treating *noninflammatory* acne. These include sulfur (2 to 8%); resorcinol (1 to 4%); benzoyl peroxide (5 to 10%); and vitamin A acid (retinoic acid, 0.025 to 0.1%). The mode of action of these preparations is not fully understood, but generally they cause desquamation (peeling of scales) and cause papules and pustules to dry and peel. Sometimes, to inhibit the growth of *P. acne*, topical antibiotics, such as tetracycline, erythromycin, and clindamycin, are prescribed.

A breakthrough in the treatment of acne vulgaris occurred in the early 1980s, when a U.S. pharmaceutical firm introduced a powerful drug called *Accutane*, which, for carefully screened and qualified patients, often can accomplish rather remarkable results in returning the patient's complexion to one that is smooth and clear. Unfortunately, however, the drug must be highly restricted because of its teratogenic effects (birth defects and deformaties) if administered to a pregnant patient or a patient contemplating pregnancy. Although not fully banned in the United States as of 1992, the drug is limited to use by experienced dermatologists and then only for extreme cases of cystic acne and *never* in situations where pregnancy may be a factor.

Additional Reading

Farmer, E. R., and A. F. Hood: *Pathology of the Skin*, Appleton and Lange, Norwalk, Connecticut, 1990.

Sams, W. M., Jr. and P. J. Lynch: *Principles and Practice of Dermatology*, Churchill Livingstone, New York, 1990.

Sauer, G. C.: *Manual of Skin Diseases*, J. B. Lippincott, Philadelphia, 1991.

Sun, M.: "Anti-Acne Drug Poses Dilemma for FDA," *Science*, 714 (May 6, 1988).

ACOELA. An order of free-living flatworms in which the alimentary tract is without a cavity.

ACOELOMATA. Animals without a coelom. The term is applied especially to the flatworms, nemertine worms, and roundworms; these animals have attained a mesoderm, in which the body cavity develops but it remains a more or less continuous mass with small internal spaces if any.

ACOUSTIC COMPENSATOR. A device for adjusting acoustical path lengths for matching purposes in binaural listening.

ACOUSTIC MODE. A type of thermal vibration of a crystal lattice which, in the limit of long wavelengths, is equivalent to an acoustic wave traveling with nearly constant velocity as if through an elastic continuum. At high frequencies, approaching the Debye frequency, the phase velocity of the acoustic modes tends to decrease, owing to dispersion. See also **Crystal.**

ACOUSTIC PRINCIPLE OF SIMILARITY. For any acoustic system involving diffraction phenomena, it is possible to construct a new system on a different scale which will perform in similar fashion, provided that the wavelength of the sound is altered in the same ratio as the linear dimensions of the original system.

ACOUSTICS. Broadly interpreted, *acoustics* is the science of sound and the technology which is associated with it.[1] Because of the complexity of the topic, it can be divided into two broad areas:

1. *Sound production and transmission*—an oscillation in pressure, stress, particle displacement, particle velocity, etc. in a medium with internal forces (for example, elastic, viscous), or the superposition of such propagated oscillations. For this definition, the terms *sound wave* or *elastic wave* or *phonon* are appropriate.

A sound wave is a longitudinal wave that consists of a sequence of pressure pulses or elastic displacements of air or other material (solid, liquid, or gas). A sound wave traveling in a gas is a sequence of compressions (dense gas) and rarefactions (less dense gas). Generally, of course, the gas in question is a mixture of nitrogen and oxygen (air). When sound travels in a solid, the sound wave is a traveling sequence of elastic compression and expansion waves. Similarly, in liquids, the sound wave is a traveling sequence of combined elastic deformation and compression waves. Propagation rate depends upon a number of variables, notably temperature, pressure, and the elastic properties of the medium. The currently accepted figure for the speed of sound in dry, open air under so-called standard conditions (0°C; 1 atm) is 331.45 meters per second (741.43 miles per hour).[2] A *phonon* is an acoustic energy packet similar in concept to a photon (light) and a function of the frequency of vibration of a sound source. Phonons are contained in or form a sound wave.

2. *Sound as perceived*—a sound sensation evoked by the oscillations previously described. Not all sound waves, of course, will evoke an auditory sensation in human receptors. Sound frequencies are usually defined: *Audible*, 20 to 20,000 Hz (cps); *higher*, ultrasonic, 20,000 to 40,000+ H_2; *lower*, infrasonic, below the audible range. For sound traveling in air at 1126 Hz, the wavelength is about one foot. In nature, the waves may be simple sinusoidal, complex, or explosive (shock), depending on the source. See also **Doppler Effect**.

Disciplines of Acoustics

Several disciplines fall within the realm of acoustics: (1) *Physical* acoustics; (2) *physiological* acoustics; (3) *psychological* acoustics, of which music is a major category, (4) *electroacoustics*, and (5) *architectural* acoustics.

Physical Acoustics

Physical acoustics deals with the fundamental *physics* and behavior of sound phenomena. Embraced are the properties and behavior of longitudinal waves of "infinetesimal" amplitude in solid, liquid, or gaseous media. These waves are propagated at the velocity of sound, or phase velocity, which is independent of frequency in a nondissipating free medium. In such a case, the shape of a complex wave remains unchanged during its propagation, although its amplitude may change. When the velocity of sound, or phase velocity, becomes dependent on frequency, the shape of a complex wave changes during propagation and dispersion is said to occur. In such cases, groups of waves comprising a limited range of frequencies travel at a velocity known as the

---

[1] IEEE Society on Acoustics, Speech, and Signal Processing.

[2] Most scientists over several decades have accepted the figure given here. S. K. Wong (National Research Council of Canada) recalculated the speed of sound from the original formula suggested by Laplace over two centuries ago. The calculation produced the figure 741.07 mph—0.05% less than the accepted speed figure. In a careful review of experiments conducted over many years, Wong found a margin of error as great or greater than that which can be ascribed to his theoretical calculation. Fortunately, for most scientific and engineering purposes, the currently accepted value serves well.

group velocity, which differs from the phase velocity. It is the group velocity which carries the energy of such complex waves.

Acoustic Waves. These waves are *dispersive* in (1) a *free medium* in which viscosity, heat conduction, and molecular, thermal, or chemical relaxation cause an increase in phase velocity with frequency; (2) in a *confined medium*, in a capillary tube, for example, in which viscosity causes a decrease in phase velocity with frequency; (3) in a *confined medium* in nondissipative tubes of increasing cross section, where the rate of change of cross-sectional area differs from the conical, i.e., different from proportionality to the square of the distance along the tube (examples of such tubes are the exponential and catenoidal horns, in which the phase velocity increases with decreasing frequency); (4) in nondissipative cylindrical tubes with *flexible* walls; (5) in waves of *finite* amplitude, where the higher-frequency components have a higher phase velocity than the lower-frequency components, a transfer of energy occurring from the lower-frequency components to the higher-frequency components.

Physical acoustics studies the reflection, refraction, diffraction, and absorption of sound waves. Properties of wave motion, such as reinforcement and destructive interference, are studied. Such waves are accompanied by pressure and particle-velocity fluctuations detectable by the ear or by instruments capable of measuring the frequency instantaneous values, and mean intensity of these fluctuations.

In *geometrical acoustics*, a subcategory of physical acoustics, phenomena are studied where diffraction and interference are disregarded. Energies of direct and reflected waves are considered to add irrespective of relative phase, a condition that applies to incoherent (i.e., uncorrelated) waves.

Physiological Acoustics

This discipline deals with animal (principally human) hearing and its impairment, the voice mechanism, and the physical effects in general of sounds on living bodies. A number of specialties exist in this field. Medical professionals are concerned with the diagnosis, treatment, and prevention of illnesses and disorders of the human sound system, i.e., the ability not only to perceive (listen) sounds, but also with the means for creating (voice) sounds. The field also includes the design and application of means for improving both the hearing and voice systems—electronic hearing aids, improvements in telephony for creating clearer more audible messages, particularly for the hard of hearing, as well as speech training for the handicapped. Much more attention over the last several years has been devoted to reducing noise pollution and the dangers of long exposure of the human ear to adverse acoustical conditions in the workplace.

Human ears do not respond, in general, to frequencies outside the audio band (20–20,000 Hz), although small animals such as cats and bats do hear in the lower ultrasonic region. At one time called supersonics, the term *ultrasonic* is now preferred to distinguish this area of high-frequency sound propagation from the noise of supersonic aircraft, supersonic fluid flow (as encountered in industrial valves), and shock waves in fluids, which have to do with speeds higher than the speeds of sound.

The *strength* of a sound field is measured by its mean square pressure expressed as sound pressure level (L_p) in decibels. Decibels are logarithmic units defining the range of sound pressure levels (L_p) between the minimum audible value at 1000 Hz (4 dB[3]—the threshold of hearing) for the average pair of good young (high school age) ears and the maximum audible value of L_p at which effects other than hearing (such as tickling in the ears—the threshold of *feeling*) begin to appear. This upper limit shows up at about 120 dB at 1000 Hz.

Higher values of L_p (e.g., 130 dB) begin to cause pain in the average ear, and values of 160 dB may well cause instantaneous physical damage (perforation) to the tympanic membrane. The minimum audible sound pressure, p_0, at 1000 Hz is internationally accepted as 0.0002 microbar rms (i.e., 0.0002 dyne/cm^2, rms), and the sound pressure level at any other rms value of sound pressure, p, irrespective of frequency, is given by $L_p = 20 \log_{10} (p/p_0)$ dB. The bel (seldom used) is simply equal to 10 decibels. The bel appears first used in connection with power loss in telephone lines and is named in honor of Alexander Graham Bell.

Other reference pressures, p_0, may be used in special applications, instead of 0.0002 microbar, so it is essential to specify the reference pressure when quoting values of L_p.

The *loudness* of a sound field is judged by the ear in the audio frequency range. Loudness judgments by groups of observers have established a *loudness level* scale. The loudness level (L_N) in phons is arbitrarily taken equal to the sound pressure level L_p in dB at the reference frequency of 1000 Hz over the range from the threshold of hearing to the threshold of feeling. Jury judgment of equality in loudness between test tones at different frequencies (f) and 1000 Hz reference tones of known sound pressure level (L_p) have established *equal loudness* contours (contours of constant L_N) in the L_p-f plane.

These contours show in general a marked decrease in ear sensitivity to sounds at frequencies below about 200 Hz, and this decrease is much more pronounced in the lower loudness levels. For example, at 50 Hz the 4-phon contour has an L_p of about 43 dB, the 80-phon contour about 93 dB. At higher frequencies, the ear shows some 8-dB increase in sensitivity in the region around 3500 Hz, then a loss in sensitivity beyond about 6000 Hz. These characteristics of hearing are significant in the design of lecture and music halls, noise-control devices, and high-fidelity audio equipment.

Also based on jury judgments, a scale of *loudness*, N, (in sones) has been established for sounds (for pure tones and for broad band noise). On this scale, a given percentage change in sone value denotes an equal percentage change in the subjective loudness of the sound. The scale provides single numbers for judging the relative loudnesses of different acoustical environments, for evaluating the percentage reduction in noise due to various noise control measures, and for setting limits on permissible noise in factories, from motor vehicles, etc.

Loudness N is related to loudness level L_N in the range 40 to 100 phons by the equation $\log_{10} N = 0.03L_N - 1.2$. A loudness of 1 sone corresponds to a loudness level of 40 phons and is typical of the low-level background noise in a quiet home.

Various methods are available for estimating loudness of complex sounds from their sound pressure levels in octave, half-octave, or third-octave bands. For traffic noises, readings on a standard sound level meter using the A-scale (which incorporates a frequency-weighting network approximating the variation of ear sensitivity with frequency to tones of 40-dB sound pressure level) appear to correlate reasonably well with jury judgments of vehicle loudness.

The *noisiness* of a broad-band noise is more related to the annoyance it causes than to its loudness. Thus, corresponding to the scale of sones created to measure loudness, a scale of *noys* has been developed as a measure of the noisiness of jet aircraft noise in particular. Noys give more importance to the high-frequency bands of noise and less importance to the low-frequency bands than do sones. Also, corresponding to the scale of loudness levels in phons, there has been established a scale of *perceived noise levels* in PN dB. Rules have been established for converting sound pressure level measured in octave bands, half-octave bands and third-octave bands into noys and then into PN dB. Although originally developed as a means for the assessment of the "noisiness" of jet aircraft flying over inhabited communities, the concept of noisiness is being applied to traffic and other broad-band noises. See also **Hearing and the Ear.**

Psychological Acoustics

This diverse discipline deals with the emotional and mental reactions of persons to various sounds. What sounds are acceptable to most people under various living conditions and what sounds are not? For what period of time can undesirable sounds be tolerated (aside from possible physical damage)? Over the years, various noise criteria have been developed that are related to the so-called preferred speech interference level PSIL. By definition, PSIL is the average of the sound pressure levels in decibels (L_p) in three octave frequency bands centered on 500 Hz, 1,000 Hz, and 2,000 Hz. The loudness level in phons (L_N) of a broadband noise (no outstanding pure tones) should not be over 22 phons (at the most, not over 30 phons) greater than the PSIL, in decibels, of the background noise.[4] Two noise control criteria (NC and NCA) are designed to fulfill these conditions for various sound fields,

---

[3] 0 dB at 1000 Hz is defined in older work as the threshold of hearing.

ranging from radio and television broadcasting studios, through bedrooms, offices, restaurants, sports arenas, and factories.

Musical Sounds. A major segment of psychological acoustics (although popularly considred more of an art than a science) is music. It is well accepted, of course, that music exerts numerous emotional responses. It is only within the last few decades that a number of researchers have investigated musical sounds from the standpoint of basic physics. A merger of electronics with sound, of course, commenced with radio, followed shortly by the extensive use of electronic equipment for amplifying and conditioning musical sounds at concerts and recording studios. Then, entered the interface between computer science and electronically composed and produced music—with the most recent impact being the digital reproduction of musical sounds on compact discs, where a laser beam serves as the equivalent of the pickup or stylus of a conventional record player.

A muscial sound may be described as an aural sensation caused by the rapid periodic motion of a sonorous body. In contrast, noise is an aural sensation due to nonperiodic motions. These observations, originally made by Helmholtz, may be modified slightly so that the frequencies of vibration of the body fall within the limits of hearing: 20 to 20,000 Hz. This definition is not clear-cut; there are some noises in the note of a harp (the twang) as well as a recognizable note in the squeak of a shoe. In other cases, it is even more difficult to make a distinction between music and noise. In some modern "electronic music," hisses and thumps are considered a part of the music. White noise is a complex sound whose frequency components are so closely spaced and so numerous that the sound ceases to have pitch. The loudness of these components is approximately the same over the whole audible range, and the noise has a hissing sound. Pink noise has its lower frequency components relatively louder than the high frequency components.

The attributes of musical sound and their subjective correlates are described briefly. The number of cycles per second, frequency, is a physical entity and may be measured objectively. Pitch, however, is a psychological phenomenon and needs a human subject to perceive it. In general, as the frequency of a sonorous body is raised, the pitch is higher. However, pitch and frequency do not bear a simple linear relationship. To show the relationship, a pitch scale can be constructed so that one note can be judged to be two times the pitch of another and so on. The unit of pitch is called the mel, and a pitch of 1,000 mels is arbitrarily assigned to a frequency of 1,000 Hz. In general, it is observed that the pitch is slightly less than the frequency at frequencies higher than 1,000 cycles, and slightly more than the frequency at frequencies less than 1,000 Hz. Pitch also depends on loudness. For a 200 cycle tone if the loudness is increased the pitch decreases, and the same happens for frequencies up to 1,000 Hz. Between 1,000 and 3,000 Hz pitch is relatively independent of loudness, while above 4,000 Hz, increasing the loudness raises the pitch. A rapid variation in pitch when the variation occurs at the rate of from two to five times per second is called vibrato. The pitch variation in mels may be large or small but the rate at which the pitch is varied is rarely greater than five times per second. Violinists produce vibrato by sliding their fingers back and forth a minute distance on a stopped string. A variation in loudness occurring at the rate of two to five times a second is called tremolo. Singers often produce a combination of tremolo and vibrato to give added color to their renditions.

Like frequency, intensity is a physical entity defined as the amount of sound energy passing through unit area per second in a direction perpendicular to the area. It is proportional to the square of the sound pressure, the latter being the rms pressure over and above the constant mean atmospheric pressure. Since sound pressure is proportional to the amplitude of a logitudinal sound wave and to the frequency of the wave, intensity is proportional to the square of the amplitude and the square of the frequency. Sound intensity is measured in watts per second per square centimeter and, since the ear is so sensitive to sound, a more usual unit is microwatt per second per square centimeter. By way of example, a soft speaking voice produces an intensity of .1 micromicrowatt/cm^2 sec, which 1,500 bass voices singing fortissimo at a distance 1 cm away produce 40 watts/cm^2 sec. Because of such large ranges of intensities, the decibel scale of intensity is normally used to designate intensity levels. An arbitrary level of 10^{-16} watts/cm^2 sec is taken as a standard for comparison at 1,000 Hz. This is very close to the threshold of audibility. At this frequency, other sound levels are compared by forming the logarithm of the ratio of the desired sound to this arbitrary one. Thus $\log I/10^{-16}$ is the number of bels a sound of intensity I has, compared to this level. Since this unit is inconveniently large, it has been subdivided into the decibel one-tenth its size; $10 \log I/10^{-16}$ equals the number of decibels (dB) the sound has. A few intensity decibel levels are listed:

	dB
Quiet whisper	10
Ordinary conversation	60
Noisy factory	90
Thunder (loud)	110
Pain threshold	120

While intensity levels can be measured physically, loudness levels are subjective and need human subjects for their evaluation. The unit of loudness is the phon, and an arbitrary level of 0 phons is the loudness of a 1,000-Hz note which has an intensity level of 0 dBB. Sounds of equal loudness, however, do not have the same intensity levels for different frequencies. From a series of experiments involving human subjects, Fletcher and Munson in 1933 constructed a set of equal loudness contours for different frequencies of pure tones. These show that for quiet sounds (a level of 5 phons) the intensity level at 1,000 cycles is about 5 dB lower than an equally loud sound at 2,000 cycles, for 30 cycles about 70 dB lower, and at 10,000 cycles about 20 dB lower. In general, as the intensity level increases, loudness levels tend to be more alike at all frequencies. This means that as a sound gets less intense at all frequencies, the ear tends to hear the higher and lower portions of sound less loudly than the middle portions. Some high fidelity systems incorporate circuitry that automatically boosts the high and low frequencies as the intensity level of the sound is decreased. This control is usually designated a loudness control.

That entity which enables a person to recognize the difference between equally loud tones of the same pitch coming from different musical instruments is called timbre, quality, or tone color. A simple fundamental law in acoustics states that the ear recognizes only those sounds due to simple harmonic motions as pure tones. A tuning fork of frequency f, when struck, causes the air to vibrate in a manner which is very nearly simple harmonic. The sound that is heard does, in fact, give the impression that it is simple and produces a pure tone of a single pitch. If one now strikes simultaneously a series of tuning forks having frequencies f (the fundamental), $2f$, $3f$, $4f$, $5f$, etc. (overtones), the pitch heard is the same as that of the single fork of frequency f except that the sound has a different quality. The quality of the sound of the series can be changed by altering the loudness of the individual forks from zero loudness to any given loudness. Another way to alter the tone quality is to vary the time it takes for a composite sound to grow and to decay. A slow growth of an envelope even though it contains the same frequencies makes for a different tone quality than one which has a rapid growth. The difference in quality between a b-flat saxophone and an oboe is almost entirely due to the difference in growth or decay time.

A fundamental theorem discovered by the mathematician Fourier states that any complicated periodic vibration may be analyzed into a set of components which has simple harmonic vibrations of single frequencies. If this method of analysis is applied to the composite tones of musical instruments, it is seen that these tones consist of a fundamental plus a series of overtones, the intensity of the overtones being different for instruments of differing timbre. Rise and decay times will also differ. The reverse of analysis is the synthesis of a musical sound. Helmholtz was able to synthesize sound by combining sets of oscillating tuning forks of various loudness to produce a single composite steady tone of a definite timbre. Modern synthesizers are more sophisticated. Electrical oscillators of the simple harmonic variety are combined electrically and then these electrical composite envelopes are electronically modified to produce differing rise and decay times. A transducer changes the electrical composite envelope into an acoustical one so that a sound of any desired timbre rise and

---

[4] A phon is the unit of loudness level. Loudness level in phons is equal to the sound pressure level in decibels with reference to 0.0002 microbar of a pure tone of 1000 Hz, which a group of listeners judge to be equally loud.

decay time can be produced. An alternate way to produce similar effects is to use an oscillation known as the square wave. When this is analyzed by the method of Fourier, it is shown to consist of a fundamental plus the odd harmonics or overtones. Another kind of oscillator, a sawtooth wave, when analyzed, is shown to consist of the fundamental and all harmonics—even and odd. A square wave or a sawtooth wave produced by an appropriate electrical oscillator can be passed through an electrical filter which can attenuate any range of frequencies of the original wave. This altered wave can later be transformed into the corresponding sound wave. In this way sounds having a desired rise and decay time, plus the required fundamental and overtone structure, can be made as desired.

Learning from Antique Instruments. In comparatively recent years, considerable interest has been shown by individuals and groups of scientists in the acoustics of early (17th and 18th century) musical instruments. The investigators use modern research tools and analytical techniques, but even with this new technology, there remains a considerable mystique as to how the early craftsmen achieved the exceptionally high sound quality of their instruments.

Hutchins (see *Additional Reading*) reports on the acoustics of violin plates. A worldwide but small organization (Catgut Acoustical Society) has used advanced scientific methods in the study of violins and other string instruments. Modern tests of the vibrational properties of the unassembled top and back plates of a violin reveal something of what the makers of violins do by intuition or "feel" in constructing consistently good violins. Complex mathematics involving the Chladni method of displaying the eignemodes of a free (unattached) violin plate and the use of hologram interferometry for determining the vibrational patterns of violin plates were central to the study reported by Hutchins.

In a scholarly paper by Richardson, the manufacture of stringed musical instruments is assessed from the standpoint of acoustics and stresses that tone quality of a stringed instrument is intimately related to the modes of vibration of its *body*. Richardson discusses the vibrations of violins and guitars and shows how mode analysis and numerical modeling may lead to a quantitative understanding of the relationships between a musical instrument's construction and its tone quality. The author stresses that the greatest difficulties arising in violin making are associated with the mechanical properties of the wood used and that the maker must compensate for such variations by modifying the dimensions of the instrument parts. Although beyond the scope of this article, the results of holographic interferometry of a guitar are shown in Fig. 1 as being illustrative of the scientific sophistication that goes into such studies.

Over scores of years, there has been much speculation pertaining to selecting the right pieces of wood (for example, by Antonio Stradivari in his violin making, or by the Ruckers family in Antwerp, who produced harpsichords with such splendid sound). There is unconfirmed speculation that early instrument makers may have inspected the grain of wood to determine its density by viewing sunlight through thin sections. Tillmann Steckner, a contemporary harpsichord maker in London, Ontario, as of 1989, had not been able to test his speculation by utilizing an x-ray technique because, "No one has given me a Stradivari to take apart."

The acoustics of the harpsichord also have been intensely studied by other investigators. One of the most intriguing aspects of this instrument is the "swirling sound" of the instrument. This is not produced by a piano. E. L. Kottick et al. have employed Chladni patterns and computer models, as have other investigators. Although the details of the research are beyond the scope of this volume, the investigators reason that the energy of the vibrating strings account for little sound directly picked up by the ear; rather, this energy is passed to the instrument's soundboard by way of the bridge. The strings actually move only a comparatively few molecules of air. Acting as a selective filter, the bridge then permits only some of the frequencies to pass to the soundboard. The latter superimposes its own characteristics on the frequencies received. Some of these are suppressed; others are enhanced. Additional factors which affect the instrument's sound include structural members and the cavity, all of which interact with the soundboard. Still another factor that contributes to the characteristics perceived by the ear is a strong bass. The researchers suggest that this is a psychophysiological phenomenon that can be likened to heterodyning, as encountered in electronics. See **Heterodyne.**

Baroque instruments other than vibrating strings also have been studied in recent yers. For example, Rossing reports of a study of the physics of kettledrums. Rossing notes that the vibrations of a violin string form a harmonic series with a distinct pitch. The vibrations of an ideal membrane, in contrast, do not form such a series. How then can a kettledrum have a pitch? The Rossing paper partially answers this question. Future research will be directed toward better understanding other percussion instruments.

A revival of baroque music is putting greater emphasis on methods of singing and playing that satisfy criteria for historically correct performances. Many of the baroque musical instruments have been successfully constructed, the surviving antique originals having yielded many of their secrets. But not so, they report, for the baroque trumpet. The instrument today is not made as it was in the 17th and 18th Centuries; consequently, playing techniques have been compromised. The investigators explain in considerable detail their studies of the instrument and generally conclude that the problem of playing the baroque trumpet resides in the contrast of how the instrument was made a few centuries ago and how modern copies are made. They conclude that only by consistently applying historical principles to all three parts of the equation—the player, the mouthpiece, and the instrument—can one accurately revive the lost art of playing the baroque trumpet.

Mathematics and Musical Sound. Harmonics, the science of musical sounds, has been taught to serious students of music and the voice for at least two centuries, and in recent years this knowledge has been expanded markedly by applying computer technology to old principles. Ironically, as pointed out by Bracewell, the human ear essentially calculates in an instant a mathematical transform when it converts sound waves of pressure traveling through the atmosphere into a sound spectrum (i.e., a series of volumes at distinct pitches). The brain further processes this information into what is known as *perceived sound*. In the late 1700s, Baron Jean-Baptiste-Joseph Fourier, a French scientist and mathematician, developed a relatively complex mathematical method for analyzing complex *fluctuating* phenomena that relate not only to sound waves but also to such apparently unrelated areas as heat conduction. Although not universally applicable, Fourier found that a sum of sinusoidal (wave-like) functions would converge to correctly represent a discontinuous function. Thus, well established for scores of years as a useful mathematical method of analyzing sound, Fourier transforms now are being applied to the study of DNA's double helix, sawtooth signals in electronics, and sunspot cycles, among others. See Fig. 2 and also **Fourier Transform.**

An example of how much more remains to be learned in the category of musical sounds is the paradoxical perception of pitch. A few decades ago, R. N. Shepard (AT&T Bell Laboratories) gathered an audience of listeners and repeatedly played an identical sequence of computer-generated tones that were moved up an octave. The panel, instead of commenting that there was undistorted repetition, curiously stated that there was a rise in pitch from one repetition to the next. Of course there was no external reason for this, and it was believed that some mechanism of human hearing was responsible for the phenomenon. Along similar lines, earlier professional musicians had hinted of the phenomenon. This puzzler of pitch perception has been termed the *tritone paradox*. Deutsch (Univ. of California, San Diego) and other researchers worldwide are exploring several hypotheses, including proposals that persons who speak different dialects of a language may perceive tonal patterns in strikingly different ways. This supports long-held speculation that the way people hear music is related in some way to their speech characteristics.

Research at MIT has included development of a computer program to assist students in writing programs that characterize certain composers. But, when students hear their programs played back, they find that such models fail in at least one critical way—i.e., some major composers, such as Bach and Vivaldi, have unique movements that to date have not yielded to formal, procedural organization.

In this present period of intensive research of musical sounds, the human voice has not been overlooked. A high-technology laboratory at the Oberlin (Ohio) College Conservatory of Music is providing new knowledge on the actions and performance of human vocal chords. In-

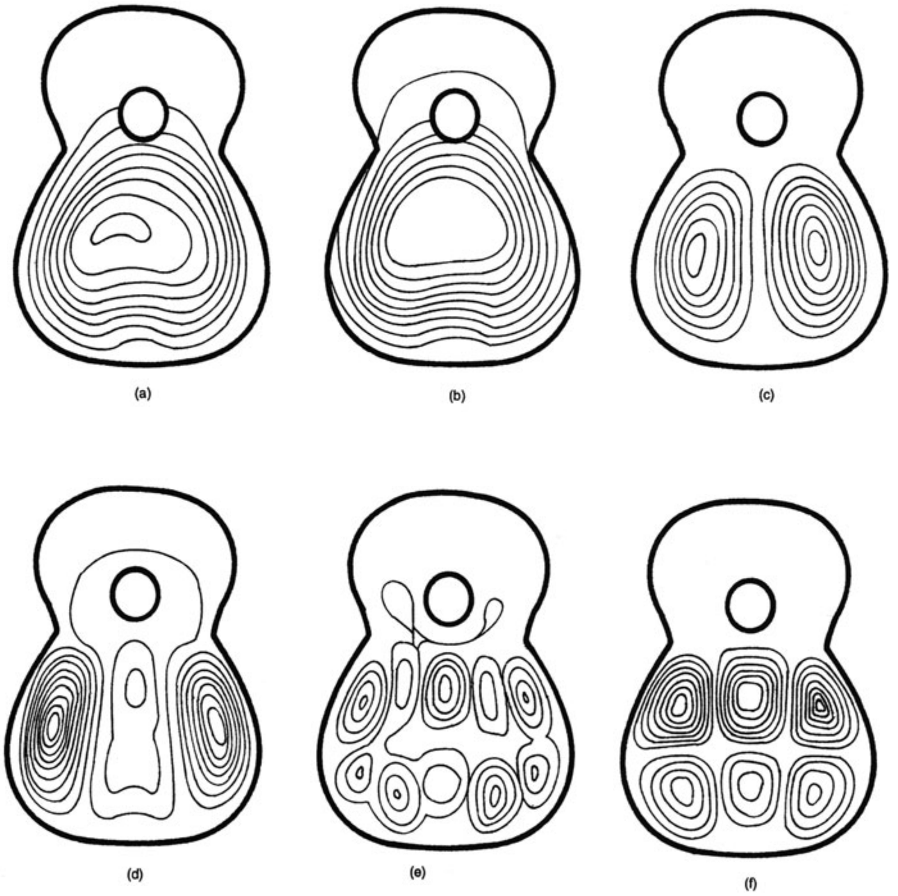


Fig. 1. To visualize the complex body vibration of musical instruments requires rather sophisticated research tools. Over many years of research, various methods have been devised to measure the fractional (tiny) displacements. These have included Chladni patterns and intricate capacitive probes. One method preferred today is that of *holographic interferometry.* This is the method used by Richardson at the University of Wales (Cardiff). Reasonable facsimiles of some of these are shown here. The modes of vibration of a guitar are visualized. Mode frequencies are (a) 106 Hz, (b) 2 to 16 Hz, (c) 268 Hz, (d) 553 Hz, (e) 1194 Hz, and (f) 509 Hz. The mode shapes and frequencies are unique to the instrument shown (guitar). The fringes shown are the result of interference between the multiple recorded by the hologram. As explained by researcher Richardson, the fringes form a contour map of the amplitude displacement of the plate (its *eigenfunction*). The system has a sensitivity on the order of 1 micrometer. These patterns reflect the construction of the particular instrument and are dependent upon such factors as the volume of the air cavity (depth of ribs and plate area) and the size of the round hole (called *F* hole in violin). The same principle that is used in bass-reflex loudspeakers, (i.e., a Helmholtz resonator) acts as the coupling between the plates and the air cavity.

strumentation used includes electronic displays of vibrato, oscillation, and tremolo. A sonograph-printer and spectrum analyzer are used to visualize a singer's resonance. An electrolaryngograph furnishes information on vocal chords, producing graphics of breathy singing. One researcher observes that the type of effort in biomechanics that has been made on behalf of sports should be duplicated in the study of vocal pedagogy. See also **Voice and Sound Production**.

The majority of scientists and musicians concerned with music research today recognize that much has been learned to date, but that the field is still in an experimental phase. Electronic music research emanates mainly from computer scientists who have strong personal interests in music. What they are seeking does not always match the interests of professional musicians, some of whom are seeking ways to enhance the effects of the human performer. This would include searching for and using new musically related algorithms. Efforts along these lines are being made at the Institut de Recherche et Coordination Acoustique/Musique (Paris), where algorithms are being created so that a computer may track live music played by adjacent instruments. One researcher has observed that "finding the pitch" is a terribly difficult problem.

In another direction, some investigators are working on the concept of a "radio baton," which would incorporate a number of transmitters that would send signals to track the movement of the baton in 3-D. This tool, in essence, could then be used to direct a computer-simulated orchestra.

The goals of numerous and separate music research laboratories remain to be fully defined and coordinated. Thus, the experimental phase probably will continue into the foreseeable future.

Digital Reproduction of Musical Sounds. Although digitization has been used to enhance recordings of great artists from the past, the major thrust of digital reproduction is that of making superior recordings of contemporary performances. The first component of audio systems to use digital processing was the phonograph, but authorities are forecasting that every element of the sound system will ultimately employ digital technology. In the present and familiar *compact disc*, the sound is preserved as a series of microscopic pits and smooth areas. A laser beam, replacing the former stylus or pickup, serves as the playback device. In the older analog type of recording, the sound consists of a continuous variation of amplitude over time. Any deviations from linearity cause distortion on the waveform. All analog systems have some nonlinearity. With analog systems, noise has been an ever-present problem. In addition to noise arising from the microphones used to make the recording, the recording medium per se is of a granular nature and introduces further noise. As pointed out by Monforte, the noise puts a lower limit on the resolving power of the storage medium. Bandwidth is another fundamental limitation of traditional recording devices.

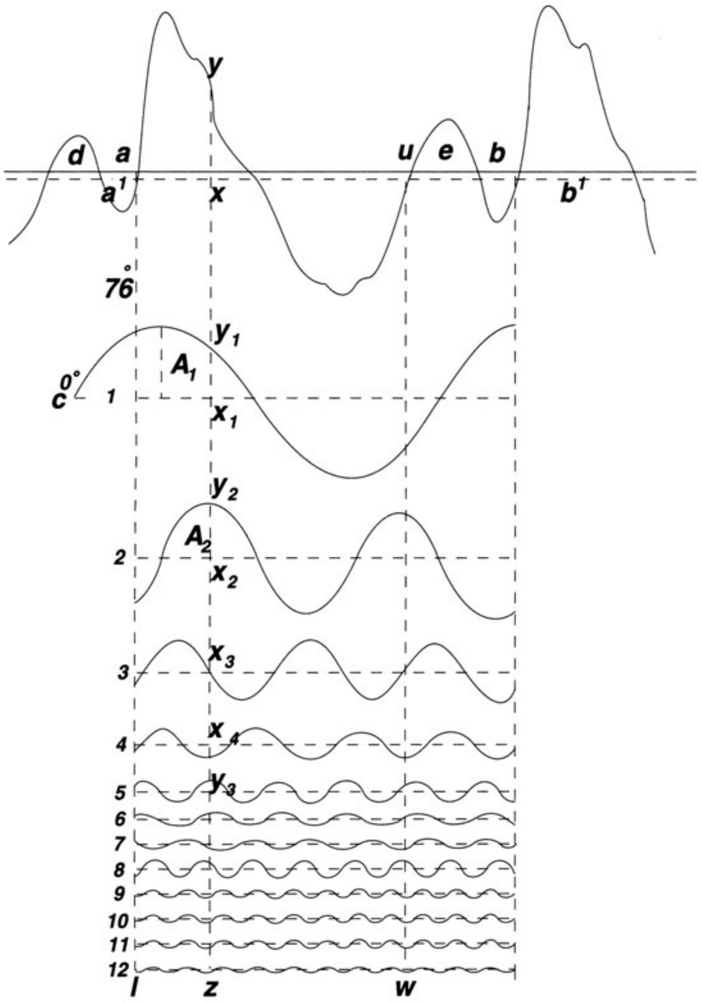


Fig. 2. Records of a complex sound and 12 of its components. Developed by Dayton C. Miller, a pioneer sound physicist (circa 1920).

Monforte describes this as a limit on how quickly the system can respond to the rapid changes in amplitude that are characteristic of sounds.

Digitization of sound dates back to the 1920s when Bell Laboratories engaged in a project to find ways to overcome the limitations of analog recording. The waveform in a digital audio system is converted to a series of numbers, which become a description of the waveform. See Fig. 3.

In digital recording, the familiar analog-to-digital (A/D) technique is used. In making this conversion, both amplitude and time must be incorporated into the digital signal. (A/D techniques have been used in the computer process control field for many years.) Reconstruction of the signal again follows earlier techniques used in other fields—the digital signal is fed to a digital-to-analog (D/A) converter. The resulting signal is a replica of the original waveform. The system is subject to two major sources of error—inadequate sampling and quantization.

Because of the very high density of stored data, the compact disc has nearly unlimited potential in electronic data processing. The compact disc was developed by the Philips Corporation of The Netherlands in cooperation with the Sony Corporation of Japan. The disc has a 500-megabyte capacity.

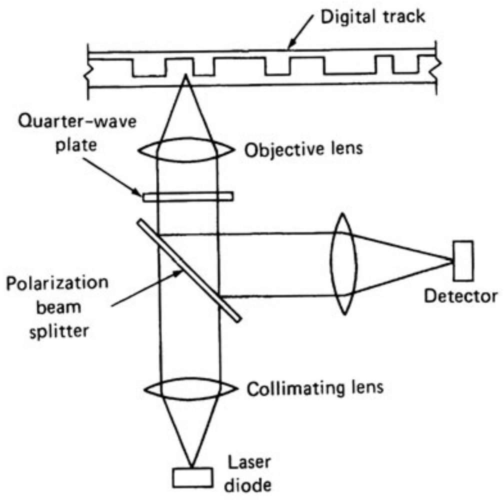


Fig. 3. Playback system used in connection with digitally recorded discs (*Compact Disc*). Light from a laser diode, shown at bottom of diagram, is passed through a collimating lens to alleviate any tendency for the beam to diverge. A polarized beam splitter divides the concentrated beam. Part of the beam is passed to the detector at right. The remaining portion of the beam passes through another filter that rotates the axis of polarization by 90 degrees. An objective lens focuses the beam on the disc surface (digital track) shown along top of diagram. Highly exaggerated bumps on the surface (about the size of the wavelength of incident light) scatter the light, preventing it from returning to the detector. Instantly, the detector senses the decline in beam intensity. These variations are read by the system as a string of binary digits (1's and 0's).

Since the late 1980s, the motion picture industry has turned, at least partially, to all-digital sound. Binary code, as compared with an analog signal on a plastic or magnetic tape, is not subject to deterioration. This is a very attractive advantage. Editing digital sound also can be simpler and less time consuming, contributing to lower cost. However, the initial switching of the film industry to all-digital technology represents a very large investment in equipment (optical storage, for example); in the meantime, the suppliers of analog sound equipment are improving their systems. Thus, the resistance to change is economic and coupled with institutional inertia. Digital recording has been gaining experience from use in shorter films and has been particularly adept for filling in special effects. Digital sound is particularly effective for recording background street noises, the clatter and rumbling of subway and rail cars, the slamming of doors, pouring liquids, and so on. Some experts still insist that, with exception of such special effects for enhancing realism, analog recording sounds the best. A larger number of experts consider the trend to digital as inevitable. However, still other experts claim that the recording method used should not be the major consideration, but that much more precise design attention should be given to improving the acoustics of theater buildings, as well as upgrading the theater's sound delivery system.

In an exceptionally interesting paper, Fletcher and Thwaites describe in detail the physics of the concert hall organ at the Sydney Opera House (Australia) which was completed in 1979. The Sydney organ, designed and built by Ronald Sharp, has some 10,500 pipes controlled by the mechanical action of five keyboards and a pedal board. The mechanical action, which regulates the flow of air into the pipes, is duplicated by an electric action that is under microprocessor control. The organ, therefore, can be operated by a magnetic tape on which an original performance has been digitally recorded.

Electroacoustics

This discipline is concerned with the principles by which electrical energy can be converted into acoustic energy and vice versa. Consider the familiar electrodynamic transducer. A periodic electric current passing through a coil interacts with a steady radial magnetic flux causing the coil to vibrate. The coil in turn drives a diaphragm which radiates sound waves from one side. (The other side is usually enclosed to avoid cancellation of the acoustic output.) The entire process is *reversible* since sound waves striking the diaphragm set up a periodic variation in air pressure adjacent to the diaphragm causing it to vibrate. As the moving coil cuts the magnetic flux, an emf is generated which causes a current to flow when a load is connected to the coil terminals.

Many, but not all, types of transducer are similarly reversible. A reversible transducer may be made to perform sending and receiving functions successively in such a manner that an absolute sensitivity may be determined (*reciprocity* calibration).

The electrodynamic transducer may further be classified as *passive* since all of the energy appearing in the acoustic load is derived from the electrical input energy, and *linear* in the sense that there is a substantially linear relationship between the input and output variables (electric current and acoustic pressure in the present case).

Irreversible Transducers. These depend on a variety of special effects of which the best known is (a) the variation of surface contact electrical resistance with pressure (carbon microphone). Other effects are (b) the variation of bulk resistance with elastic strain (piezoresistance), (c) variation of transistor parameters with strain, (d) cooling effect of periodic air movement (hot wire microphone), (e) pressure wave generated by an electrical spark, (f) dependence of air pressure on level of corona discharge (ionophone). See also **Microphone.**

Reversible Transducers. An important class of reversible transducer depends on relative movement of suitable components linked by an electric or magnetic field traversing a gap. Examples are (a) the electrodynamic transducer already described; (b) electrostatic depending on the relative movement of charged condenser plates; (c) magnetic or variable reluctance depending on relative movement of magnetic poles in a magnetic circuit linked with a fixed coil.

Other *reversible transducers* are dependent on dimensional changes connected with the state of magnetic or electric polarization of certain crystalline materials (piezomagnetism and piezoelectricity). Since strain may be longitudinal or shear and since both strain and polarization are directional quantities, many possible relationships between strain and polarization exist. The behavior of an X-cut quartz disk may serve as an illustration. When such a disk is axially compressed, electric charges appear on the plane surfaces. Conversely, if a potential difference is established between the two surfaces, contraction or expansion occurs depending on the direction of the electric field. Other important single-crystal piezoelectric materials are ammonium dihydrogen phosphate (ADP) and Rochelle salt. During the past decade, polycrystalline ceramic materials based on barium titanate and lead zirconate titanate have replaced single-crystal materials in many applications. These materials are ferroelectric and, when prepolarized, exhibit piezoelectric behavior.

To date, only polycrystalline piezomagnetic materials (often termed magnetostrictive) have been found useful. Some are metals such as nickel and permendur. Others are ferrite ceramics [basic composition: $(NiO)(Fe_2O_3)$] which have such a high electrical resistivity that eddy current losses are negligible making lamination unnecessary.

Electromechanical Coupling. Transducer performance is closely connected with the tightness of coupling between mechanical and electrical aspects. Consider a piezoelectric disk which is compressed by putting in *mechanical* energy W_m. The appearance of surface charges shows that *electrical energy* W_e is stored in the self capacitance and is available when an external circuit is connected to suitable electrodes. The ratio W_e/W_m (electromechanical coupling coefficient) sets a limit to the efficiency for a given bandwidth (frequency range). The coefficient may reach 70% for lead zirconate titanate.

Transducer Design. Impedance matching is of primary importance in electroacoustics. It may be likened to the choice of gear ratio and wheel size in automobile design. Impedance matching is generally closely related to transducer parameters such as beam width of projected or received sound and frequency response, as well as efficiency. The many available matching technique include (a) Resonance, (b) horn systems (acoustic transformers), (c) lever systems (mechanical transformers). In the direct radiator electrodynamic loudspeaker, the diaphragm is made large enough to interact with the acoustic medium (air) and yet small enough in relation to the sound wavelength (at low frequencies, at least) to ensure uniform projection of sound over a wide angle. In the condenser loudspeaker, a large transducer area compensates for the weakness of electrostatic forces. In the underwater sonar project, slabs of piezoelectric ceramic may be sandwiched between metal plates to form a resonant device which radiates a narrow beam of sound with high efficiency over a narrow frequency range.

Since 1970, much progress was made toward refining an acoustic (voice) interface between people and computers. This topic is explored in the entry on **Telephony;** and **Voice and Sound Production,** among others. In the development of systems and components for various kinds of voice communications, it is necessary in the performance of various tests to isolate a chamber (anechoic chamber) as much as possible from ambient radiation, including sound and other electromagnetic radiation. A chamber of this type is shown in Fig. 4.

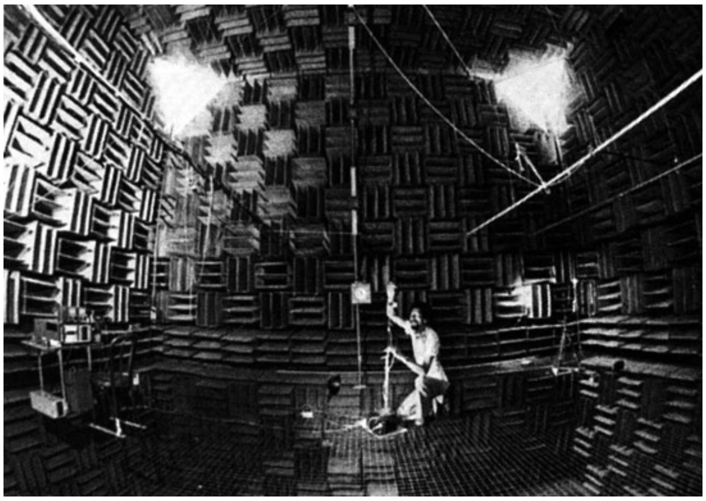


Fig. 4. Anechoic chamber, a superquiet space for testing acoustic components and systems. Adjustments on a directional microphone are being made in preparation for determining its directional characteristics. (*AT&T Bell Laboratories.*)

Microwave Acoustics. This field is concerned with the use of acoustic waves in solids for signal storage, amplification, and processing in the frequency range above 50 MHz. A piezoelectric transducer thin enough to operate in the fundamental mode at several hundred MHz can be formed by evaporating a thin film of piezoelectric material onto a suitable substrate or by forming a semiconductor *depletion layer* of the correct thickness. *Magnetoelastic transducers* have been formed of materials such as yttrium iron garnet (YIG) which operate in the resonance mode of the ferromagnetic spin system and generate longitudinal acoustic waves.

A sound-transmitting bar and a pair of transducers provides an effective and compact *delay line*. The interaction of free charges with elastic waves in piezoelectric materials can provide acoustic amplification, which can be achieved with a longitudinal electric field sufficient to establish a carrier drift velocity greater than the elastic wave velocity. *Traveling-wave amplifiers* with 40 dB gain and 10% bandwidth at 1 GHz for a one millimeter transmission path have been built with semiconductors.

The *interdigital transducer* is an array of parallel conducting strips with $\lambda/2$ spacing deposited on a piezoelectric substrate, such as lithium niobate, which provides efficient excitation of acoustic surface waves (Rayleigh waves). Surface waves generated in this fashion can be guided and selectively delayed by grooves and metallic film boundaries and can be coupled in and out at many points along the path. Surface waves also can be amplified by drifting charge carriers in the substrate or in a semiconductor layer above the wave-carrying surface. These properties are compatible with integrated circuit techniques.

The *electret microphone* uses an electrostatic transducer in which a polarizing field is maintained by a quasi-permanent charge layer embedded in a thin plastic film. The electret transducer has a very high electrical impedance and can be combined with an integrated field effect transistor (FET) amplifier. Complex array properties can be built into electret microphones. Examples include the second-order gradient (toroidal) microphone for conference use and a square array for acoustic holography.

The *parametric acoustic array* has provided a means for obtaining a narrow beam of low-frequency underwater sound, using a small primary transducer. Due to the nonlinearity of the equations of fluid motion, a pair of highly collimated high-frequency sound beams can be made to act as a very large end-fire array, launching a directional sound beam at a comparatively low difference frequency. Since the liquid medium rapidly absorbs the primary beams, the array is *tapered*.

Intense coherent sound waves can be generated at several GHz by the electrostrictive processes which accompany the passage of intense laser beams through liquids and solids (stimulated Brillouin scattering).

Architectural Acoustics

In this field, scientists and engineers deal with the problems of distribution of beneficial sounds within buildings and with the exclusion of reduction of undesirable sounds. It has been known for many years that the mass and limpness of barriers, such as partitions, are highly important in providing high sound transmission loss. In auditoriums, reflective ceilings and reflective walls, combined with convex irregularities of random design, provide for reinforcement and diffuseness of sound found so beneficial for speech and music. Reflecting surfaces, giving short time-delay reflections (about 20 ms or less), are particularly desirable in concert halls. Delays of 65 ms or more may result in echoes and speech unintelligibility.

Publications of specifications for environmental and architectural acoustics are obtainable from the American Society for Testing and Materials.

As of the early 1990s, major advancements are occurring in architectural acoustics. An example is a computer graphics program developed at Cornell University, in which a wireframe rendition of a symphony hall is constructed. As the first hall to be studied, the researchers selected the design of Boston Symphony Hall, which is famous for its superior acoustics. A simulated sound is created from center stage. Initially, this expands in the form of a simple sphere in bold colors on the screen. Reflections are then produced from the ceiling and balconies of the hall, but in subtler hues of color. As the signal dies out, only a few uncolored areas remain, thus identifying locations in the hall without sound. When designing a new hall or contemplating design improvements, a special wireframe representative of the geometry of the target hall can be made. One of the researchers on this project observes that each sound can be observed as it travels all over the hall and in steps of 1 ms, which is a finer resolution than what the human ear can perceive. With just a few instructions to the system, one can select the best from a series of possible hall design situations, such as the best ceiling slopes, balcony arrangements, ceiling and inside wall building materials, and so forth.

This technique may eliminate the need for constructing intricate physical models that normally use light waves instead of sound as the testing medium.

Automotive Acoustics

Traditionally, predicting noise in the passenger compartment of an automobile in the early design stage was considered unachievable. Acoustic analysis was simple—build a prototype vehicle, get in, listen, then try to calm the cacophony. This trial-and-error acoustic engineering created numerous problems, for once a noise was designed in, it was difficult and costly to eliminate it in the prototype or production models. Vehicle body structures are much like metal drums. Vibrations induced by the road, by the car's aerodynamic loadings, and by the power train all generate sound in much the same way as the vibrating skin of a drum.

Several years ago, General Motors (U.S.) developed a computer-based acoustic model by mapping out the complex dimensional geometry of the passenger vehicle interior. The model was a useful analytical tool because it could define booming frequencies and anticipate design problems. But the model did not provide sufficient information to identify which part of the structure created the noise, nor, more importantly, did it indicate what modifications would reduce the noise.

The next step was development of a system in which the acoustic model was coupled with a computer-aided structural model. The structural model could simulate vibration responses of body surfaces during actual vehicle operations. With a coupling equation, the structural model could be linked with the acoustic model to translate vibration data into the sound effects. The combined structural-acoustic model can identify noise paths, pinpointing the parts of the body structure which are likely to create noise and the percentage which various panels and modes contribute. The model can suggest structural modifications, such as redesigning a panel, early in the design stage to avoid costly add-on solutions which also may add weight to the vehicle. The new methodology has been verified by a comparison study between the structural-acoustic model's predictions and actual prototype vans driven over rough roads at the GM proving grounds. As shown by Fig. 5, the model accurately predicted both acoustic peaks and overall noise levels. A new systems approach to acoustic prediction and modification is shown in diagram form in Fig. 6.

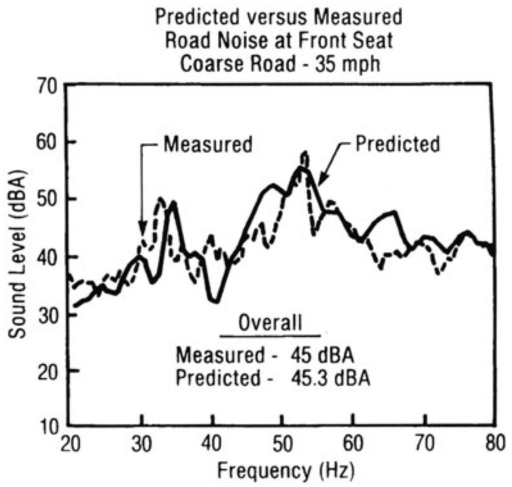


Fig. 5. Comparison of the acoustic computer model's predictions made in the laboratory versus actual measurements made on the proving ground. Tests were for a prototype van and show the close relationship between predictions and actual noise experience. (*General Motors Corp.*)

Other Applications of Acoustic Phenomena

Ultrasound in Chemistry. Chemical reactions between two or more materials involve energy transfer. A reaction may (1) spontaneously generate energy (most often thermal), or (2) require the absorption of external energy to proceed partially or fully to completion. In the latter case, ultrasound frequently is an excellent source of external, additional energy. Generally, the application of acoustic energy to chemistry is termed *sonochemistry*.

Ultrasound in high-energy chemistry is particularly effective when one or more liquid reactants are involved. Mixtures of homogeneous liquids or slurries (liquid-solids systems) are good examples. The application of ultrasound normally is successful because of a phenomenon known

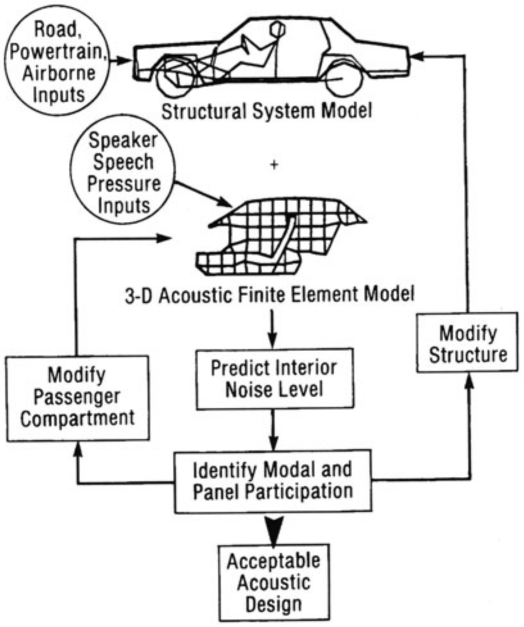


Fig. 6. Systems approach to acoustic prediction and modification for producing more acceptable vehicle designs during the earliest, pre-prototype stages in the development of a new vehicle. System predicts what parts of the car or van structure that create noise and how this noise is transmitted. (*General Motors Corp.*)

as *acoustic cavitation*. See also **Cavitation**. The latter occurs when ultrasound creates bubbles in the reactant(s), which expand rapidly, followed by what is known as *implosive collapse,* during which localized spots of very high temperature (5000°C; 9000+°F) are created and accompanied by very high pressures (up to 500 atmospheres). This process occurs within a few microseconds, creating shock waves that travel at extremely high velocities, causing, in a liquid-solid slurry, for example, impacts (collisions) between particles. It is estimated that the strength of such energy releases could instantaneously melt most metals. In fact, ultrasound is commonly used to clean metal surfaces that are highly reactive, as well as for increasing the effectiveness of catalytic reactions.

Ultrasound, with certain materials, also can cause the emission of light. Known as *sonoluminescence*, this property was observed by Frenzel and Schultes (1934). Although not applied to practical advantage, it has been found in recent years to apply to nonaqueous as well as water components. Currently, sonoluminescence is considered to be a form of *chemiluminescence*. The relatively recent emergence of ultrasound's importance in chemistry is very well summarized in the Suslick (1990) reference listed.

Instrumentation and Testing. Acoustic emissions, particularly ultrasound, are widely and variously used for testing the properties of materials and for imaging (medical applications being just one example). In recent years, ultrasonic microscopy has evolved from a laboratory to an industrial inspection technique. The use of acoustic emissions for testing pressure vessels used in industry have been commonplace for many years. The use of ultrasonic nondestructive characterization of materials has been widely used for years. These and other applications are described in various parts of this *Encyclopedia*. Check alphabetical index.

Additional Reading

Bamberger, J.: "Computerizing Vivaldi," *Technology Rev. (MIT)*, **13** (July 12, 1990).

Botermans, J., Dewit, H., and H. Goddefroy: "Making and Playing Musical Instruments," Univ. of Washington Press, Seattle, Washington, 1989.

Bracewell, R. N.: "The Fourier Transform and Its Applications," 2nd Edition, McGraw-Hill, New York, 1986.

Bracewell, R. N.: "The Fourier Transform," *Sci. Amer.*, 86 (June 1989).

Corcoran, E.: "Sound Bytes—Electronic Music Gains a Human Touch," *Sci. Amer.*, 111 (July 1991).

Deutsch, D.: "Auditory Pattern Recognition," in "Handbook of Perception and Human Performance" (R. Boff, et al., Editors) Wiley, New York, 1986.

Deutsch, D.: "The Tritone Paradox: An Influence of Language on Music Perception," *Music Perception*, **8** (4) 335 (Summer 1991).

Deutsch, D.: "Some New Pitch Paradoxes and Their Implications," *Philosophical Trans. of the Royal Socy. of London*, Series B, **336** (1278) 391 (June 1992).

Deutsch, D.: "Paradoxes of Musical Pitch," *Sci. Amer.*, 88 (August 1992).

Dloktycz, S. J., and K. S. Suslick: "Interparticle Collisions Driven by Ultrasound," *Science*, 1067 (March 2, 1990).

Eldridge, W.: "Synthetic Sound, Real People," *Technology Rev. (MIT)*, 76 (August/September 1991).

Fletcher, N. H., and S. Thwaites: "The Physics of Organ Pipes," *Sci. Amer.*, 94 (January 1983).

Fletcher, N. H., and T. D. Rossing: "The Physics of Musical Instruments," Springer-Verlag, New York, 1991.

Gordon, S.: "The New Music Biz," *Technology Rev. (MIT)*, **10** (January 10, 1987).

Graff, G.: "Virtual Acoustics Puts Sound in its Place," *Science*, 616 (May 1, 1992).

Horgan, J.: "Stradivari's Secret," *Sci. Amer.*, 21 (July 1989).

Kottick, E. L., Marshall, K. D., and T. J. Hendrickson: "The Acoustics of the Harpsichord," *Sci. Amer.*, 110 (February 1991).

Longreth, R. N.: "Mother Tongue May Influence Musical Ear," *Sci. News*, 138 (December 1, 1990).

Moffat, A. S.: "New Graphics Program Debuts in Concert Hall," *Science*, 1452 (September 29, 1989).

Parker, S. P., Editor: "Acoustics Source Book," McGraw-Hill, New York, 1988.

Richardson, B. E.: "Vibrations of Stringed Musical Instruments, *Review (Univ. of Wales)*, 13 (Autumn 1988).

Shepard, R. N.: "Circularity in Judgments of Relative Pitch," *J. Acoustical Socy. Amer.*, **36** (12) 2346 (December 1964).

Smithers, D., Woegram, K., and J. Bowsher: "Playing the Baroque Trumpet," *Sci. Amer.*, 108 (April 1986).

Staff: "Music Science," *Technology Rev. (MIT)*, 72 (January 1990).

Strauss, S.: "Software Soprano," *Technology Rev. (MIT)*, 12 (July 1988).

Sturges, D.: "Sounding Out Ceramic Quality," *Adv. Mat. & Processes*, 35 (April 1991).

Suslick, K. S.: "The Chemical Effects of Ultrasound," *Sci. Amer.*, 80 (February 1989).

Suslick, K. S., Editor: "Ultrasound—Its Chemical, Physical, and Biological Effects," VCH, New York, 1988.

Suslick, K. S.: "Sonochemistry," *Science*, 1439 (March 23, 1990).

Wade, A. P., et al.: "An Analytical Perspective on Acoustic Emission," *Analytical Chemistry*, 497A (May 1, 1991).

Wright, K.: "Digital Audio Is Changing Film Sound—Will Audiences Notice?" *Sci. Amer.*, 35 (June 1989).

ACOUSTIC SCINTILLATION. Irregular fluctuations in the received intensity of sounds propagated through the atmosphere from a source of uniform output. These variations are produced by the nonhomogenous structure of the atmosphere along the path of sound. Turbulence and its concomitant variations in temperature and moisture are the chief causes of the inhomogeneities that lead to sonic refraction, diffraction, and scattering responsible for acoustic scintillation.

ACOUSTICS (Telephone). See **Telephony.**

ACOUSTICS (Underwater). See **Radar.**

ACROMEGALY. A disease associated with the pituitary gland. The condition has been known since antiquity, but was not specifically described as a distinct clinical syndrome until 1886 by P. Marie, a French physician and medical researcher. The pituitary source of acromegaly was confirmed by Cushing in 1909. Dr. Cushing proposed that the cause is an excessive secretion of growth-promoting hormone by a hyperfunctioning pituitary gland. Within the last decade or two, much progress has been made toward a better understanding of acromegaly. This is well documented by Melmed (1990). When the underlying conditions develop early in life, the result may be *giantism*, where individuals may attain a height of 8 feet (2.4 meters) or more. If the condition develops later in life, after bones have ceased to grow, an overactive pituitary causes excessive stimulation of growth centers, resulting in the condition known as *acromegaly*. This condition is frequently characterized by an abnormal development of feet and hands, an abnormally prominent jaw, and enlarged bones of the skull, sometimes causing the appearance like that of a primitive human.

Acromegaly occurs just as frequently in both men and women; the annual incidence is estimated at 3 to 4 cases per million and thus is rare. Mean age at diagnosis is about 40 years in men and 45 years in women. It is important to point out, however, that, because of so many unreported or undiagnosed cases, the occurrence worldwide is from 50 to 70 cases per million of population. Major life-threatening diseases associated with acromegaly include hypertension, diabetes, pulmonary infections, and cancer. Because the mortality associated with acromegaly is approximately double that of healthy subjects at the same age, once diagnosed it should be treated aggressively. The underlying cause is a pituitary adenoma, which normally shows up by using magnetic resonance imaging or computerized tomography. Once identified, therapy includes surgical treatment, radiation treatment, and drug therapy, including bromocriptine, a lysergic-acid–ergot derivative and a dopamine agonist. A recently developed long-acting somatostatin analogue octreotide also appears promising.

See also **Pituitary Gland**.

Additional Reading

Becker, K. L.: "Principles and Practice of Endocrinology and Metabolism," J. B. Lippincott, Philadelphia (1990).

Lloyd, R. V.: "Endocrine Pathology," Springer-Verlag, New York, 1990.

Melmed, S.: "Acromegaly," *N. Eng. J. Med.*, 966 (April 5, 1990). Note: This reference contains an excellent bibliography.

Moore, W. T., and R. C. Eastman, Editors: "Diagnostic Endocrinology," B. C. Decker, Philadephia, 1989.

ACRYLIC ACID. $CH_2{:}CH{\cdot}COOH$, formula weight 72.06, colorless liquid monocarboxylic acid, mp 12°C, bp 141°C, sp gr 1.062. Also called propenoic acid, this compound is miscible in all proportions with H_2O or alcohol. The acid forms esters and metallic salts and forms addition products. The compound is of particular interest because of the large number of synthetic plastics and resins which are made as the result of polymerizing various acrylic derivatives, notably the esters of acrylic acid. The anhydrous monomer, glacial acrylic acid, contains less than 2% H_2O. It yields esters when reacted with alcohols, including ethyl acrylate and methyl acrylate. See also **ABS (Acrylonitrile-Butadiene-Styrene) Resins; Acrylonitrile;** and **Fibers.**

ACRYLIC PLASTICS. A wide range of plastic materials dates back to the pioneering work of Redtenbacher before 1850 who prepared acrylic acid by oxidizing acrolein

$$CH{=}CHCHO \xrightarrow{O} CH_2{=}CHCOOH$$

At a considerably later date, Frankland prepared ethyl methacrylate and methacrylic acid from ethyl α-hydroxyisobutyrate and phosphorus trichloride. Tollen prepared acrylate esters from 2,3-dibromopropionate esters and zinc. Otto Rohm, in 1901, described the structures of the liquid condensation products (including dimers and trimers) obtained from the action of sodium alkoxides on methyl and ethyl acrylate. Shortly after World War I, Rohm introduced a new acrylate synthesis, noting that an acrylate is formed in good yield from heating ethylene cyanohydrin and sulfuric acid and alcohol. A major incentive for the development of a clear, tough plastic acrylate was for use in the manufacture of safety glass.

Ethyl methacrylate went into commercial production in 1933. The synthesis proceeded in the following steps:

(1) Acetone and hydrogen cyanide, generated from sodium cyanide and acid, gave acetone cyanohydrin

$$HCN + CH_3COCH_3 \rightarrow (CH_3)_2C(OH)CN$$

(2) The acetone cyanohydrin was converted to ethyl α-hydroxyisobutyrate by reaction with ethyl alcohol and dilute sulfuric acid

$$(CH_3)_2C(OH)CN + C_2H_5OH \xrightarrow{H_2SO_4} (CH_3)_2C(OH)COOC_2H_5$$

(3) The hydroxy ester was dehydrated with phosphorus pentoxide to produce ethyl methacrylate

$$(CH_3)_2C(OH)COOC_2H_5 \xrightarrow{P_2O_5} CH_2{=}C(CH_3)COOC_2H_5$$

In 1936, the methyl ester of methacrylic acid was introduced and used to produce an "organic glass" by cast polymerization. Methyl methacrylate was made initially through methyl α-hydroxyisobutyrate by the same process previously indicated for the ethyl ester. Over the years, numerous process changes have taken place and costs lowered, making these plastics available on a very high tonnage basis for thousands of uses. For example, the hydrogen cyanide required is now produced catalytically from natural gas, ammonia, and air.

As with most synthetic plastic materials, they commence with the monomers. Any of the common processes, including bulk, solution, emulsion, or suspension systems, may be used in the free-radical polymerization or copolymerization of acrylic monomers. The molecular weight and physical properties of the products may be varied over a wide range by proper selection of acrylic monomer and monomer mixes, type of process, and process conditions.

In **bulk polymerization** no solvents are employed and the monomer acts as the solvent and continuous phase in which the process is carried out. Commercial bulk processes for acrylic polymers are used mainly in the production of sheets, rods and tubes. Bulk processes are also used on a much smaller scale in the preparation of dentures and novelty items and in the preservation of biological specimens. Acrylic castings are produced by pouring monomers or partially polymerized sirups into suitably designed molds and completing the polymerization. Acrylic bulk polymers consist essentially of poly(methyl methacrylate) or copolymers with methyl methacrylate as the major component. Free radical initiators soluble in the monomer, such as benzoyl peroxide, are the catalysts for the polymerization. Aromatic tertiary amines, such as dimethylaniline, may be used as accelerators in conjunction with the peroxide to permit curing at room temperature. However, colorless products cannot be obtained with amine accelerators because of the formation of red or yellow colors. As the polymerization proceeds, a considerable reduction in volume occurs which must be taken into consideration in the design of molds. At 25°C, the shrinkage of methyl methacrylate in the formation of the homopolymer is 21%.

Solutions of acrylic polymers and copolymers find wide use as thermoplastic coatings and impregnating fluids, adhesives, laminating materials, and cements. Solutions of interpolymers convertible to thermosetting compositions can also be prepared by inclusion of monomers bearing reactive functional groups which are capable of further reaction with appropriate crosslinking agents to give three-dimensional polymer networks. These polymer systems may be used in automotive coatings and appliance enamels, and as binders for paper, textiles, and glass or nonwoven fabrics. Despite the relatively low molecular weight of the polymers obtained in solution, such products are often the most appropriate for the foregoing uses. Solution polymerization of acrylic esters is usually carried out in large stainless steel, nickel, or glass-lined cylindrical kettles, designed to withstand at least 50 psig. The usual reaction mixture is a 40–60% solution of the monomers in solvent. Acrylic polymers are soluble in aromatic hydrocarbons and chlorohydrocarbons.

Acrylic **emulsion polymers and copolymers** have found wide acceptance in many fields, including sizes, finishes and binders for textiles, coatings and impregnants for paper and leather, thermoplastic and thermosetting protective coatings, floor finishing materials, adhesives, high-impact plastics, elastomers for gaskets, and impregnants for asphalt and concrete.

Advantages of emulsion polymerization are rapidity and production of high-moleuclar-weight polymers in a system of relatively low viscosity. Difficulties in agitation, heat transfer, and transfer of materials are minimized. The handling of hazardous solvents is eliminated. The two principal variations in technique used for emulsion polymerization are the redox and the reflux methods.

Suspension polymerization also is used. When acrylic monomers or their mixtures with other monomers are polymerized while suspended (usually in aqueous system), the polymeric product is obtained in the form of small beads, sometimes called pearls or granules. Bead polymers are the basis of the production of molding powders and denture materials. Polymers derived from acrylic or methacrylic acid furnish exchange resins of the carboxylic acid type. Solutions in organic solvents furnish lacquers, coatings and cements, while water-soluble hydrolysates are used as thickeners, adhesives, and sizes.

The basic difference between suspension and emulsion processes lies in the site of the polymerization, since initiators insoluble in water are used in the suspension process. Suspensions are produced by vigorous and continuous agitation of the monomer and solvent phases. The size of the drop will be determined by the rate of agitation, the interfacial tension, and the presence of impurities and minor constituents of the recipe. If agitation is stopped, the droplets coalesce into a monomer layer. The water serves as a dispersion medium and heat-transfer agent to remove the heat of polymerization. The process and resulting product can be influenced by the addition of colloidal suspending agents, thickeners, and salts.

Product Groupings. The principal acrylic plastics are cast sheet, molding powder, and high-impact molding powder. The cast acrylic sheet is formable, transparent, stable, and strong. Representative uses include architectural panels, aircraft glazing, skylights, lighted outdoor signs, models, product prototypes, and novelties. Molding powders are used in the mass production of numerous intricate shapes, such as automotive lights, lighting fixture lenses, and instrument dials and control panels for autos, aircraft, and appliances. The high-impact acrylic molding powder yields a somewhat less transparent product, but possesses unusual toughness for such applications as toys, business machine components, blow-molded bottles, and outboard motor shrouds. The various acrylic resins find numerous uses as previously mentioned, with varied and wide use in coatings. Acrylic latexes are composed mainly of monomers of the acrylic family, such as methyl methacrylate, butyl methacrylate, methyl acrylate, and 2-ethyl hexylacrylate. Additional monomers, such as styrene or acrylonitrile, can be polymerized with acrylic monomers. Acrylic latexes vary considerably in their properties, mainly affected by the monomers used, the particle size, and the surfactant system of the latex. Generally, acrylic latexes are cured by loss of water only, do not yellow, possess good exterior durability, are tough, and usually have good abrasion resistance. The acrylic polymers are reasonably costly and some latexes do not have very good color compatibility. Acrylic latex paints can be used for concrete floors, interior flat and semigloss finishes, and exterior surfaces. See also **Paint and Coatings.**

A developing and potentially large volume use for acrylics is in the video, audio, and data-storage disk markets. The properties, as well as ease of fabrication, have made acrylics a primary choice for these applications. More detail on acrylics can be found in an article by J. W. Altman, *Modern Plastics Encyclopedia*, 13–16, *Modern Plastics* magazine, New York (1986–1987).

ACRYLONITRILE. $CH_2{:}CHCN$, formula weight 29.04, liquid, bp 78°C. Also called vinyl cyanide or propene nitrile, this compound is a very high-tonnage chemical used as an intermediate in the production of acrylonitrile-based plastics, nitrile rubbers, acrylic fibers, insecticides, and numerous other synthetic materials. Manufacturing processes use propylene, NH_3, and air as raw materials in what may be termed an ammonoxidation or oxyamination reaction:

$$CH_3CH{:}CH_2 + NH_3 + 1\text{-}1/2O_2 \rightarrow CH_2{:}CHCN + 3H_2O.$$

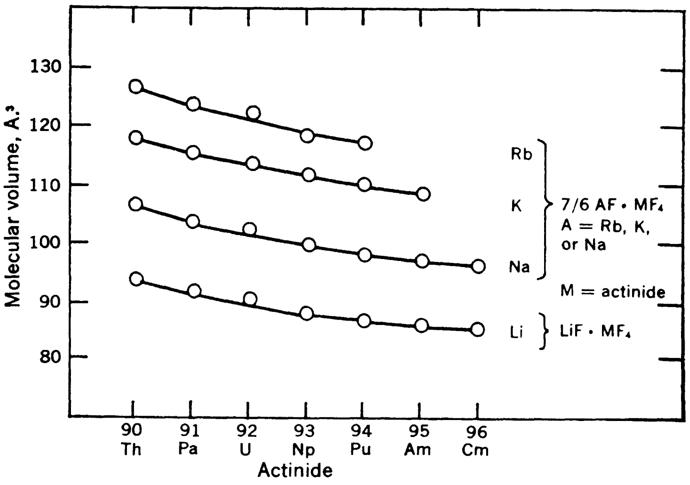
In one process, the starting ingredients are mixed with steam, preheated, and fed to the reactor. There are two main by-products, acetonitrile ($CH_3 \cdot CN$) and HCN, with accompanying formation of small quantities of acrolein, acetone, and acetaldehyde. The acrylonitrile is separated from the other materials in a series of fractionation and absorption operations. A number of catalysts have been used, including phosphorus, molybdenum, bismuth, antimony, tin, and cobalt.

ACTH. The adrenocorticotropic hormone of the anterior lobe of the pituitary gland, which specifically stimulates the adrenal cortex to secrete cortisone, and hence has effects identical with those of cortisone. ACTH differs in its chemistry, absorption, and metabolism from the other adrenal steroids. Chemically, it is a water-soluble polypeptide having a molecular weight of about 3000. Its complete amino acid sequence has been determined. It produces its peripheral physiological effects by causing discharge of the adrenocortical steroids into the circulation. ACTH has been extracted from pituitary glands. In purified form, ACTH is useful in treating some forms of arthritis, lupus erythematosus, and severe skin disorders. The action of ACTH injections parallels the result of large quantities of naturally formed cortisone if they were released naturally. See also **Adrenal Glands; Hormones; Nervous System and The Brain; Pituitary Gland;** and **Steroids.**

ACTIN. One of the two proteins that make up the myofibrils of striated muscles. The other protein is myosin. The combination of these two proteins is sometimes spoken of as actinomyosin. The banded nature of the myofibrils is due to the fact that both proteins are present where the bands are dark and only one or the other is present in the light bands. Since these bands lie side by side in the different myofibrils that go to make up a muscle fiber, the entire muscle fiber shows a banded or striated appearance. See also **Contractility and Contractile Proteins.**

ACTINIDE CONTRACTION. An effect analogous to the Lanthanide contraction, which has been found in certain elements of the Actinide series. Those elements from thorium (atomic number 90) to curium (atomic number 96) exhibit a decreasing molecular volume in certain compounds, such as those which the actinide tetrafluorides form with alkali metal fluorides, plotted in the accompanying diagram. The effect here is due to the decreasing crystal radius of the tetrapositive actinide ions as the atomic number increases. Note that in the Actinides the tetravalent ions are compared instead of the trivalent ones as in the case of the Lanthanides, in which the trivalent state is by far the most common.

The behavior is attributed to the entrance of added electrons into an (inner) *f* shell (4*f* for the Lanthanides, 5*f* for the Actinides) so that the increment they produce in atomic volume is less than the reduction due to the greater nuclear charge.



Plot of molecular volume versus atomic number of the tetravalent Actinides.

ACTINIDE SERIES. The chemical elements with atomic numbers 90 to 103, inclusively, commencing with 90 (thorium) and through 103 (lawrencium) frequently are termed, collectively, the Actinide Series. The term derives from actinium (at. no. 89) which is considered the anchor element of the series, also appearing in group 3 of the periodic table. Members of the series are listed in the accompanying table. Some authorities place actinium in the series per se. This series of elements is somewhat analogous to the Lanthanide Series. See also **Lanthanide Series.**

Justification for the grouping is found in the higher elements of (III) oxidation states similar to actinium, and (IV) oxidation states similar to thorium. Certain similarities also exist between the atomic spectra and magnetic properties in the Actinide and Lanthanide Series. Note that actinium ($Z = 89$) and thorium ($Z = 90$) differ in electronic configuration from their immediate predecessor in atomic number, radium ($Z = 88$) in having, respectively, 1 and 2 electrons in their 6*d* subshells. The next element, protactinium ($Z = 91$), is the first to have an electron of the 5*f* subshell. Note also that the configurations of the last seven elements as given in the table are enclosed in parentheses to indicate that they are predicted, rather than determined, configurations. Two major methods have been used in making these determinations for the first eight elements: emission spectroscopy for actinium, thorium, uranium and americium; and atomic-beam experiments for protactinium, neptunium, plutonium and curium.

While in many respects the electronic configurations and chemical properties of the Actinide elements are similar to those of the Lanthanide series, the 4*f*, 5*d* and 6*s* subshells of the latter corresponding to the 5*f*, 6*d*, and 7*s* of the former, there are, however, significant differences. Cerium in the Lanthanide Series, unlike its analog thorium in the Actinide Series, has an electron in its 4*f* subshell. Moreover, for the first few members of each series, the 5*f* and 6*d* electrons are less energetically bound to the atomic nucleus than the 4*f* and 5*d* ones, so that the first few Actinide elements (except actinium) have in general higher oxidation states (lose electrons more readily) than the corresponding Lanthanides. Thus uranium, neptunium, plutonium, and americium have all four of the oxidation states 3, 4, 5 and 6. Later in the series, the Actinides correspond more closely to the Lanthanides in this respect.

In their electronic configurations the Actinide elements all have their innermost 86 electrons arranged in the configuration of radon, and their additional electrons as shown in the accompanying table.

As noted in a paper by T. J. Marks (see references), the stoichiometric and catalytic chemistry of metal-organic compounds having actinide-to-carbon bonds is in a stage of new interest and growth. Chemical, structural, and bonding characteristics have been identified which differ in several ways from those of *d*-block transition element compounds. In his highly informative article, much too detailed to review here, Marks addresses such topics as pi-bonded ligands, actinide-carbon sigma bonds and their synthesis, the chemical characteristics of actinide-carbon sigma bonds, hydride formation through actinide-carbon bond hydrogenolysis, and carbon monoxide activation. There is much interest in understanding the processes by which carbon mon-

ELECTRONIC CONFIGURATIONS FOR NEUTRAL ATOMS OF THE ACTINIDE ELEMENTS

Element	Atomic Number (Z)	Electronic Configuration
Actinium	89	$6d7s^2$
Thorium	90	$6d^27s^2$
Protactinium	91	$5f^26d7s^2$
Uranium	92	$5f^36d7s^2$
Neptunium	93	$5f^46d7s^2$
Plutonium	94	$5f^67s^2$
Americium	95	$5f^77s^2$
Curium	96	$5f^76d7s^2$
Berkelium	97	($5f^86d7s^2$ or $5f^97s^2$)
Californium	98	($5f^{10}7s^2$)
Einsteinium	99	($5f^{11}7s^2$)
Fermium	100	($5f^{12}7s^2$)
Mendelevium	101	($5f^{13}7s^2$)
Nobelium	102	($5f^{14}7s^2$)
Lawrencium	103	($5f^{14}6d7s^2$)

oxide (as probably derived from coal in the future) can be catalytically transformed into several useful organic chemicals. In his conclusions, Marks observes, "Organoactinide chemistry is entering a period of rapid development. It is apparent that a rich and diverse chemistry is emerging, and that 'tuning' of the ligation sphere and f-electron configuration can exert significant control over the rate and course of many ususual chemical transformations. To place this chemistry in perspective, it appears that there are distinct similarities to main group and transition metal chemistry, but there are also pronounced differences. It is the exploitation of these latter characteristics that offers the greatest challenge and promise."

Additional Reading

Hammond, C. R.: "The Elements" in *Handbook of Chemistry and Physics*, 67th edition, CRC Press, Boca Raton, Florida (1986–1987).

Katz, J. J., and G. T. Seabord: "The Chemistry of the Actinide Elements," Methuen, London, 1957.

Marks, T. J.: "Actinide Organometallic Chemistry," *Science*, **217**, 989–997 (1982).

Staff: *Handbook of Chemistry and Physics*, 73rd, Ed. CRC Press, Boca Raton, Florida, 1992–1993.

ACTINIUM. Chemical element symbol Ac, at. no. 89, at. wt. 227 (mass number of the most stable isotope), periodic table group 3, classed in the periodic system as a higher homologue of lanthanum. The electronic configuration for actinium is

$$1s^22s^22p^63s^23p^63d^{10}4s^24p^64d^{10}4f^{14}5s^25p^65d^{10}6s^26p^66d^17s^2.$$

The ionic radius (Ac^{+3}) is 1.11Å.

Presently, 24 isotopes of actinium, with mass numbers ranging from 207 to 230, have been identified. All are radioactive. One year after the discovery of polonium and radium by the Curies, A. Debierne found an unidentified radioactive substance in the residue after treatment of pitchblende. Debierne named the new material *actinium* after the Greek word for ray. F. Giesel, independently in 1902, also found a radioactive material in the rare-earth extracts of pitchblende. He named this material *emanium*. In 1904, Debierne and Giesel compared the results of their experimentation and established the identical behavior of the two substances. Until formulation of the law of radioactive displacement by Fajans and Soddy about ten years later, however, actinium definitely could not be classed in the periodic system as a higher homologue of lanthanum.

The isotope discovered by Debierne and also noted by Giesel was ^{227}Ac which has a half-life of 21.7 years. The isotope results from the decay of ^{235}U (AcU-*actinouranium*) and is present in natural uranium to the extent of approximately 0.715%. The proportion of Ac/U in uranium ores is estimated to be approximately 2.10^{-10} at radioactive equilibrium. O. Hahn established the existence of a second isotope of actinium in nature, ^{228}Ac, in 1908. This isotope is a product of thorium decay and logically also is referred to as *meso*-thorium, with a half-life of 6.13 hours. The proportion of mesothorium to thorium ($MsTh_2$/ Th) in thorium ores is about 5.10^{-14}. The other isotopes of actinium were found experimentally as the result of bombarding thorium targets. The half-life of 10 days of ^{225}Ac is the longest of the artificially-produced isotopes. Although occurring in nature as a member of the neptunium family, ^{225}Ac is present in extremely small quantities and thus is very difficult to detect.

^{227}Ac can be extracted from uranium ores where present to the extent of 0.2 mg/ton of uranium and it is the only isotope that is obtainable on a macroscopic scale and that is reasonably stable. Because of the difficulties of separating ^{227}Ac from uranium ores, in which it accompanies the rare earths and with which it is very similar chemically, fractional crystallization or precipitation of relevant compounds no longer is practiced. Easier separations of actinium from lanthanum may be effected through the use of ion-exchange methods. A cationic resin and elution, mainly with a solution of ammonium citrate or ammonium-α-hydroxyisobutyrate, are used. To avoid the problems attendant with the treatment of ores, ^{227}Ac now is generally obtained on a gram-scale by the transmutation of radium by neutron irradiation in the core of a nuclear reactor. Formation of actinium occurs by the following process:

$$^{226}\mathrm{Ra}(\mathrm{n},\gamma)^{227}\mathrm{Ra} \xrightarrow{\beta-} {}^{227}\mathrm{Ac}$$

In connection with this method, the cross section for the capture of thermal neutrons by radium is 23 barns (23×10^{-24} cm^2). Thus, prolonged radiation must be avoided because the accumulation of actinium is limited by the reaction ($\sigma = 500$ barns):

$$^{227}\mathrm{Ac}(\mathrm{n},\gamma)^{228}\mathrm{Ac}(\mathrm{MsTh}_2)\rightarrow{}^{228}\mathrm{Th}(\mathrm{RdTh})$$

In 1947, F. Hageman produced 1 mg actinium by this process and, for the first time, isolated a pure compound of the element. It has been found that when 25 g of $RaCO_3$ (radium carbonate) are irradiated at a flux of 2.6×10^{14} ncm $^{-2}$s^{-1} for a period of 13 days, approximately 108 mg of ^{227}Ac (8 Ci) and 13 mg of ^{228}Th (11 Ci) will be yielded. In an intensive research program by the Centre d'Etude de l'Energie Nucléaire Belge, Union Minière, carried out in 1970–1971, more than 10 g of actinium were produced. The process is difficult for at least two reasons: (1) the irradiated products are highly radioactive, and (2) radon gas, resulting from the disintegration of radium, is evolved. The methods followed in Belgium for the separation of ^{226}Ra, ^{227}Ac, and ^{228}Th involved the precipitation of $Ra(NO_3)_2$ (radium nitrate) from concentrated HNO_3, after which followed the elimination of thorium by adsorption on a mineral ion exchanger (zirconium phosphate) which withstand high levels of radiation without decomposition.

Metallic actinium cannot be obtained by electrolytic means because it is too electropositive. It has been prepared on a milligram-scale through the reduction of actinium fluoride in a vacuum with lithium vapor at about 350°C. The metal is silvery white, faintly emits a blue-tinted light which is visible in darkness because of its radioactivity. The metal takes the form of a face-centered cubic lattice and has a melting point of 1050 ± 50°C. By extrapolation, it is estimated that the metal boils at about 3300°C. An amalgam of metallic actinium may be prepared by electrolysis on a mercury cathode, or by the action of a lithium amalgam on an actinium citrate solution (pH = 1.7 to 6.8).

In chemical behavior, actinium acts even more basic than lanthanum (the most basic element of the lanthanide series). The mineral salts of actinium are extracted with difficulty from their aqueous solutions by means of an organic solvent. Thus, they generally are extracted as chelates with trifluoroacetone or diethylhexylphosphoric acid. The water-insoluble salts of actinium follow those of lanthanum, namely, the carbonate, fluoride, fluosilicate, oxalate, phosphate, double sulfate of potassium. With exception of the black sulfide, all actinium compounds are white and form colorless solutions. The crystalline compounds are isomorphic.

In addition to its close resemblance to lanthanum, actinium also is analgous to curium (Z = 96) and lawrencium (Z = 103), both of the group of trivalent transuranium elements. This analogy led G. T. Seaborg to postulate the actinide theory, wherein actinium begins a new series of rare earths which are characterized by the filling of the 5*f* inner electron shell, just as the filling of the 4*f* electron shell characterizes the Lanthanide series of elements. However, the first elements of the Actinide series differ markedly from those of actinium. Notably, there is a multiplicity of valences for which there is no equivalent among the lanthanides. See **Chemical Elements** for other properties of actinium.

Mainly, actinium has been of interest from a scientific standpoint. However ^{227}Ac has been proposed as a source of heat in space vehicles. It is interesting to note that the heat produced from the absorption of the radiation emitted by 1 g of actinium, when in equilibrium with its daughters, is 12,500 cal/hour.

See also **Actinide Series**.

Additional Reading

Bouissières. G.: "Actinium dans le nouveau traité de chimie minérale de P. Pascal," pp. 7 and 1413–1446. Masson, Paris, 1960.

Centre d'Étude de l'Énergie Nucléaire Belge, Union Minière: Programme Actinium, Annual Report, 1969, Brussels.

Hageman, F.: The Chemistry of Actinium, in G. T. Seaborg and J. J. Katz (editors), "The Actinide Elements," National Nuclear Energy Series, IV-14A, p. 14, McGraw-Hill, New York, 1954.

Hammond, C. R.: "The Elements" in *Handbook of Chemistry and Physics*, 67th Edition, CRC Press, Boca Raton, Florida (1986–1987).

Katz, J. J., and G. T. Seaborg: "The Chemistry of the Actinide Elements," pp. 5–15, Methuen, London, 1957.

Kirby, H. W.: "The Analytical Chemistry of Actinium," Mound Laboratory, Miamisburg, Ohio, 1966.

Marks, T. J.: "Actinide Organometallic Chemistry," *Science*, **217**, 989–997 (1982).

Staff: "Handbook of Chemistry and Physics," 73rd Ed., CRC Press, Boca Raton, Florida, 1992–1993.

ACTINOLITE. The term for a calcium-iron-magnesium amphibole, the formula being $Ca_2(Mg,Fe)_5Si_8O_{22}(OH)_2$ but the amount of iron varies considerably. It occurs as bladed crystals or in fibrous or granular masses. Its hardness is 5–6, sp gr 3–3.2, color green to grayish green, transparent to opaque, luster vitreous to silky or waxy. Iron in the ferrous state is believed to be the cause of its green color. Actinolite derives its name from the frequent radiated groups of crystals. Essentially it is an iron-rich tremolite; the division between the two minerals is quite arbitrary, with color the macroscopic definitive factor—white for tremolite, green for actinolite. Actinolite is found in schists, often with serpentine, and in igneous rocks, probably as the result of the alteration of pyroxene. The schists of the Swiss Alps carry actinolite. It is also found in Austria, Saxony, Norway, Japan, and Canada in the provinces of Quebec and Ontario. In the United States actinolite occurs in Massachusetts, Pennsylvania, Maryland, and as a zinc-manganese bearing variety in New Jersey. See also **Amphibole; Tremolite;** and **Uralite**.

ACTINOMETER. The general name for any instrument used to measure the intensity of radiant energy. In earlier usage, the term was often restricted to the measurement of photochemically active radiation, but is now used more generally.

In meteorological and astronomical applications, actinometers may be classified according to the type of radiation they measure: (1) The *pyrheliometer* measures the intensity of direct solar radiation. It consists of a radiation sensing element enclosed in a casing that has a small aperture through which the direct solar rays enter, and a recorder unit. The amount of radiant energy absorbed is determined from the temperature rise of the sensing element. (2) The *pyranometer* measures the combined intensity of direct solar radiation and diffuse sky radiation, i.e., radiation reaching the earth's surface after having been scattered from the direct solar beam by molecules or suspensoids in the atmosphere. It consists of a recorder and a radiation sensing element mounted so that it views the entire sky. (3) The *pyrgeometer* measures the effective terrestrial radiation, i.e., the difference between the total outgoing infrared radiation of the earth's surface and the downcoming atmospheric radiation. It consists of four manganin strips, two blackened and two polished. The blackened strips are allowed to radiate to the atmosphere, whereas the polished strips are shielded. The electrical power required to equalize the temperature of the four strips is taken as a measure of the outgoing radiation.

ACTINOMYCOSIS. An infectious disease caused by Gram-positive bacteria with a characteristic filamentous branching shape which are known as actinomycetes. Because of their unusual shape, the actinomycetes were once thought to be fungi. However, they differ from fungi in their cell wall composition, notably in the lack of chitin, and their filaments or hyphae exceed one micrometer in diameter tending to fragment into bacillary forms of subculture. There are two genera of *Actinomycetacea: Nocardia* and *Actinomycetes.* The majority of human infections are, however, caused by one species, *A. israeli*, which is an aerobic commensal of the oropharynx, normally resident in gingivodental crevices and tonsillar crypts.

Actinomycosis is an endogenous infection and case to case transmission is only likely to occur through human bites or bare knuckle injuries by the opponent's teeth. It is a subacute or chronic granulomatous disease which progresses to fibrosis and suppuration with formation of external sinuses. *A. bovis* causes a condition known as "lumpy jaw" in cattle, but it rarely, if ever, causes human disease although the distinctive term has been applied also to cervicofacial actinomycosis.

In the past, the majority of cases of actinomycosis occurred in men, and cervicofacial, thoracic, and abdominal forms of the disease formerly constituted 95% of cases. Currently, female pelvic actinomycosis is probably the most common form of the disease.

The types of actinomycosis are generally classified on the basis of anatomical location. *Cervicofacial* actinomycosis (lumpy jaw) is associated with dental caries and usually follows orofacial trauma, dental extraction or gingival infection, together with poor dental hygiene. *Pulmonary* actinomycosis may follow aspiration of the organism from the oropharynx or by infiltration through the esophageal wall, giving rise to cough, pleuritic symptoms, fever, and weight loss. *Gastrointestinal* actinomycosis is acquired by ingestion of the organism which penetrates the intestinal wall at a site where the mucosa has been injured by disease or traumatized.

Symptoms most frequently are encountered in the right iliac fossa where differential diagnosis from appendicitis, Crohn's disease, or tuberculosis is important. *Hepatic* actinomycosis may account for up to 15% of cases, occurring primarily or secondary to pre-existing abdominal disease. The kidney may also become involved by local extension or through bacteremic spread. *Pelvic* actinomycosis may originate from the upward spread of organisms which reach the perineum from the oropharynx and via the intestinal tract. Pelvic extension from the abdomen may also occur, but the disease mostly has been observed in women fitted with plastic intrauterine contraceptive devices, notably the Dalkon shield. The extension of infection ranges from cervicitis and endometritis to salpingitis and tubovarian abscess.

Wherever the site of the infection, however, actinomycosis characteristically produces a chronic granulomatous inflammation which burrows deeply through contiguous tissue without respect to fascial planes. Consequently, draining sinus tracts are often present. The drainage may be purulent or sanguineous. If present, sulfur granules, small white-to-yellow aggregates of bacterial filaments, are important clues to diagnosis. The disease follows an indolent but progressive course and, in the typical case, symptoms are present for one to five months before diagnosis.

The diagnosis of actinomycosis depends upon the microscopical identification of the organism after smear or culture. Although wide surgical excision is not ordinarily required for treatment, some surgical debridement may be helpful therapeutically. Otherwise, antibiotics are the mainstay of treatment, with penicillin being the drug of choice. In the penicillin-allergic patient, tetracycline is preferred. With vigorous antibiotic treatment, the prognosis is excellent.

Additional Reading

Brown, J. R.: "Human Actinomycosis," *Human Pathology*, **4**, 319 (1973).

Mandell, G. L., Douglas, R. G. Jr., and J. E. Bennett, Editors: *Principles and Practice of Infectious Diseases*, Churchill Livingstone, New York, 1990.

Mitchell, R. G.: "Actinomycosis," in *Oxford Textbook of Medicine* (Weatherall, Ledingham and Warrell, Eds.), Oxford Univ. Press, 1983.

Monte, De la, et al.: "Systemic Actinomycosis Infection," *J. Amer. Med. Assn.*, **248**. 1876 (1982).

Robboy, S. J., and A. Vickery: "Distinguishing Actinomycosis and Nocardosis," *New Eng. J. Med.*, **282**, 593 (1970).

Scully, R. E., et al., Editors: "Actinomyces israelii (Case 29-1990)," *N. Eng. J. Med.*, 183 (July 19, 1990).

Scully, R. E., et al., Editors: "Pelvic Actinomycosis (Case 10-1992)," *N. Eng. J. Med.*, 692 (March 5, 1992).

Ann C. Vickery, Ph.D., Assoc. Prof., College of Public Health,
University of South Florida, Tampa, Florida.

ACTINON. The name of the isotope of radon (emanation), which occurs in the naturally-occurring actinium series (see **Chemical Elements; Radioactivity**), being produced by alpha-decay of actinium X, which is itself a radium isotope. Actinon has an atomic number of 86, a mass number of 219, and a half-life of 3.92 seconds, emitting an alpha particle to form polonium-215 (Actinium A).

ACTION. The action of a dynamical system is the space integral of the total momentum of the system. Specifically, if $\mathbf{r}_j$ is the position vector of the jth particle of the system, $\dot{\mathbf{r}}_j$ is its time rate of change, and m_j is the mass, the action for the path going from P_1 to P_2 is

$$\int_{P1}^{P2} \sum m_j \dot{\mathbf{r}}_j \cdot d\mathbf{r}_j$$

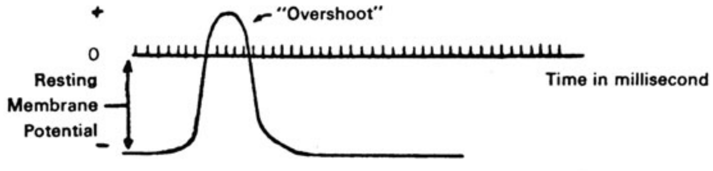
where the integral is taken along the actual path from P_1 to P_2. The integral can be shown to reduce to the form

$$2\int_{t1}^{t2} E_K \, dt$$

where E_K is the total kinetic energy of the system and t_1 and t_2 are the times at which the system is in positions P_1 and P_2 respectively. See **Least Action (Principle of).**

ACTION CURRENT. The local flow of current into the depolarized region of the cell membrane during the generation of the action potential. Since the flow of current from adjacent regions is an outward current—namely, a depolarizing current—under normal conditions it serves to depolarize the cell membrane in such regions beyond the threshold. By repetition of this process in successive regions, a self-propagated depolarization of the membrane sweeps along the nerve fiber. The nerve impulse is thus a consequence of the local action currents.

ACTION POTENTIAL. A characteristic variation in the membrane potential (i.e., potential across cell membrane) of excitable cells when the cell is stimulated. The potential falls rapidly in time toward zero, "overshoots," making the inside positive for a brief interval of time, and finally returns to the resting state. In many nerve and muscle cells, the duration of the action potential is found to be fractions of a millisecond to several milliseconds (see figure). The action potential is triggered by a depolarizing current when the membrane potential falls to the threshold level. See **Action Current.**



Action potential.

The action polarization is of particular interest in the passage of an impulse along a nerve fiber. At rest, the interior of a nerve fiber is negative to the exterior, with a potential difference of 50 to 100 millivolts; stimulation of the nerve sets up a negative wave of 70 to 100 millivolts, lasting one millisecond, followed by a brief refractory period during which further stimulation is ineffective. The record of such a change, as shown by a sensitive galvanometer, exhibits a characteristic sharp spike. Somewhat similar changes accompany muscular contraction. The study of action potentials has much practical importance e.g., electrocardiography, electroencephalography, and electromyography. See also **Nervous System and The Brain.**

ACTION SPECTRUM. A graph of the amount of biological response produced by incident light as a function of wavelength. Thus, one might construct a dose-response curve for each wavelength of light used in the inactivation of an enzyme by ultraviolet light. The relative efficiency with which the inactivation is produced plotted versus wavelength of the incident light is the action spectrum for the inactivation of that enzyme.

ACTIVATED CARBON. See **Adsorption Operations; Decolorizing Agent.**

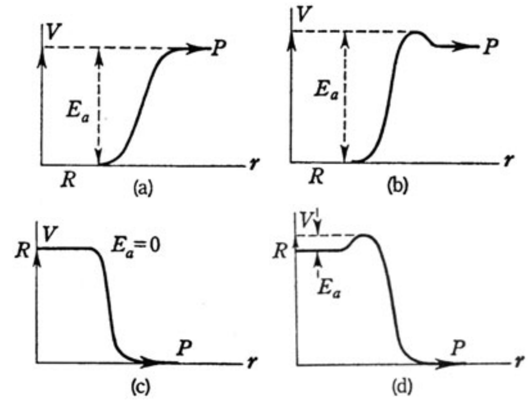
ACTIVATION. 1. The transformation of any material into a more reactive form, or into a form in which it functions more effectively, as in the regeneration of a metallic or inorganic catalyst, the transformation of an enzyme from inactive form to active form, and the treatment of various forms of finely-divided silica or carbon to render them more adsorbent.

2. The transfer of a sufficient quantity of energy to an atomic or molecular system to raise it to an excited state in which it can participate in a process not possible when the system is in its ground state.

3. In nuclear physics, the process of inducing radioactivity through neutron bombardment or by other types of radiation.

ACTIVATION ENERGY. 1. The excess energy over the ground state which must be acquired by an atomic or molecular system in order that a particular process may occur. Examples are the energy needed by the molecule to take part in a chemical reaction, by an electron to reach the conduction band in a semiconductor, and by a lattice defect to move to a neighboring site.

In the first example cited, the rate of an elementary chemical reaction can usually be expressed as a product of a function of the concentrations of the participants and of a rate constant. This latter can be written as $A \exp(-E_a/kT)$, where k is the Boltzmann constant; T, the absolute temperature; A, a frequency factor that varies slowly with the activation energy E_a, which appears in the exponential. It is the minimum height of the potential barrier that must be crossed when one follows the reaction coordinate from the reactants to the products. The figure illustrates four typical situations. It is to be noted that it is only in case (a) that E_a is equal to the energy of the reaction.



Activation energy. Variation of the energy along the reaction coordinate for two endothermic (a) and (b) and two exothermic (c) and (d) reactions. E_a is the activation energy.

2. If a liquid is regarded as an imperfect solid, the yielding to an applied shear stress takes place at a rate that depends on the frequency with which molecules leave their positions in the imperfect crystal lattice. The variation of this frequency with temperature is described by the energy required for an interchange of a molecule between the lattice and the free volume in the liquid. If this activation energy is linearly dependent on temperature at constant pressure, the slope of the $\log_e \eta : 1/kT$ plot gives the activation energy for liquid flow at low temperatures, where η is viscosity; k, the Boltzmann constant; and T, the absolute temperature.

ACTIVATOR. A substance that renders a material or a system reactive; commonly, a catalyst. 2. A special use of this term occurs in the flotation process, where an activator assists the action of the collector. 3. An impurity atom, present in a solid, that makes possible the effects of luminescence, or markedly increases their efficiency. Examples are copper in zinc sulfide, and thallium in potassium chloride. See also **Enzyme.**

ACTIVE CENTER. Atoms which by their position on a surface, such as at the apex of a peak, at a step on the surface or a kink in a step, or on the edge or corner of a crystal, share with neighboring atoms an abnormally small portion of their electrostatic field and, therefore, have a large residual field available for catalytic activity or for adsorption.

ACTIVE DEPOSIT. The name given to the radioactive material that is deposited on the surface of any substance placed in the neighborhood of a preparation containing any of the naturally occurring radioactive chains (uranium, thorium, or actinium chains). This deposit results from deposition of the nongaseous products of the gaseous radon nuclides that have escaped from the parent substance. An active deposit can be concentrated on a negatively charged metal wire or surface placed in closed vessels containing the radon. See also **Radioactivity.**

ACTIVE MASS. Mass per unit volume, usually expressed in moles per liter (a concentration factor).

ACTIVE TRANSPORT (Cell). See **Cell (Biology).**

ACTIVITY COEFFICIENT. A fractional number which when multiplied by the molar concentration of a substance in solution yields the chemical activity. This term provides an approximation of how much interaction exists between molecules at higher concentrations. Activity coefficients and activities are most commonly obtained from measurements of vapor-pressure lowering, freezing-point depression, boiling-point elevation, solubility, and electromotive force. In certain cases, activity coefficients can be estimated theoretically. As commonly used, activity is a relative quantity having unit value in some chosen standard state. Thus, the standard state of unit activity for water, a_w, in aqueous solutions of potassium chloride is pure liquid water at one atmosphere pressure and the given temperature. The standard state for the activity of a solute like potassium chloride is often so defined as to make the ratio of the activity to the concentration of solute approach unity as the concentration decreases to zero.

ACTIVITY (Radioactivity). The activity of a quantity of radioactive nuclide is defined by the ICRU as $\Delta N/\Delta t$, where N is the number of nuclear transformations that occur in this quantity in time Δt. The symbol Δ preceding the letters N and t denotes that these letters represent quantities that can be deduced only from multiple measurements that involve averaging procedures. The special unit of activity is the curie, defined as exactly $3.7 \text{ X } 10^{10}$ transformations per second. See **Radioactivity.**

ACTIVITY SERIES. Also referred to as the *electromotive series* or the *displacement series*, this is an arrangement of the metals (other elements can be included) in the order of their tendency to react with water and acids, so that each metal displaces from solution those below it in the series and is displaced by those above it. See accompanying table. Since the electrode potential of a metal in equilibrium with a solution of its ions cannot be measured directly, the values in the activity series are, in each case, the difference between the electrode potential of the given metal (or element) in equilibrium with a solution of its ions, and that of hydrogen in equilibrium with a solution of its ions. Thus in the table, it will be noted that hydrogen has a value of 0.000. In experimental procedure, the hydrogen electrode is used as the standard with which the electrode potentials of other substances are compared. The theory of displacement plays a major role in electrochemistry and corrosion engineering. See also **Corrosion;** and **Electrochemistry.**

STANDARD ELECTRODE POTENTIALS (25°C)

Reaction		Volts
$Li^+ + e^-$	$\rightleftharpoons Li$	−3.045
$K^+ + e^-$	$\rightleftharpoons K$	−2.924
$Ba^{2+} + 2e^-$	$\rightleftharpoons Ba$	−2.90
$Ca^{2+} + 2e^-$	$\rightleftharpoons Ca$	−2.76
$Na^+ + e^-$	$\rightleftharpoons Na$	−2.711
$Mg^{2+} + 2e^-$	$\rightleftharpoons Mg$	−2.375
$Al^{3+} + 3e^-$	$\rightleftharpoons Al$	−1.706
$2H_2O + 2e^-$	$\rightleftharpoons H_2 + 2OH^-$	−0.828
$Zn^{2+} + 2e^-$	$\rightleftharpoons Zn$	−0.763
$Cr^{3+} + 3e^-$	$\rightleftharpoons Cr$	−0.744
$Fe^{2+} + 2e^-$	$\rightleftharpoons Fe$	−0.41
$Cd^{2+} + 2e^-$	$\rightleftharpoons Cd$	−0.403
$Ni^{2+} + 2e^-$	$\rightleftharpoons Ni$	−0.23
$Sn^{2+} + 2e^-$	$\rightleftharpoons Sn$	−0.136
$Pb^{2+} + 2e^-$	$\rightleftharpoons Pb$	−0.127
$2H^+ + 2e^-$	$\rightleftharpoons H_2$	0.000
$Cu^{2+} + 2e^-$	$\rightleftharpoons Cu$	+0.34
$I_2 + 2e^-$	$\rightleftharpoons 21^-$	+0.535
$Fe^{3+} + e^-$	$\rightleftharpoons Fe^{2+}$	+0.77
$Ag^+ + e^-$	$\rightleftharpoons Ag$	+0.799
$Hg^{2+} + 2e^-$	$\rightleftharpoons Hg$	+0.851
$Br_2 + 2e^-$	$\rightleftharpoons 2Br$	+1.065
$O_2 + 4H^+ + 4e^-$	$\rightleftharpoons 2H_2O$	+1.229
$Cr_2O_7^{2-} + 14H^+ + 6e^-$	$\rightleftharpoons 2Cr^{3+} + 7H_2O$	+1.33
$Cl_2(gas) + 2e^-$	$\rightleftharpoons 2Cl$	+1.358
$Au^{3+} + 3e^-$	$\rightleftharpoons Au$	+1.42
$MnO_4^- + 8H^+ + 5e^-$	$\rightleftharpoons Mn^{2+} + 4H_2O$	+1.491
$F_2 + 2e^-$	$\rightleftharpoons 2F^-$	+2.85

ACTUATOR (Control System). An actuator is that portion of a final control element in a control system that furnishes the power to change and/or to maintain the valve plug position in response to a signal received from the automatic controller. In some applications, the actuator may position elements other than a valve, such as a louver, damper, or pump speed governor. Actuators may be grouped into several categories: (1) mechanical, (2) pneumatic, (3) hydraulic, and (4) electric. The handwheel on a valve is the simplest form of mechanical actuator. A simple spring-return diaphragm actuator is an example of a pneumatic actuator. Oil-operated cylinders are hydraulic actuators. Motor-operated and solenoid actuators fall into the electrical category. Combination forms include electrohydraulic and electropneumatic actuation. Actuators generally mount directly on the valve body. See **Final Control Element; Valve (Control).**

The foregoing types of actuators are those generally used in the automation of fluid (gas and liquid) processes common in the petroleum, chemical, and allied manufacturing industries. Although also used to some extent in the discrete-piece manufacturing industries, such as metals fabrication and assembly (as typically found in the automotive, aircraft, electrical, and mechanical equipment manufacturing industries, among others), the final controlling actuators frequently take other forms, such as robots and mechanized placement devices. See **Automation.**

ACUTE (Medical). Having a rapid onset, severe symptoms, and a relatively short duration; not chronic.

ACYL. An organic radical of the general formula, RCO—. These radicals are also called acid radicals, because they are often produced from organic acids by loss of a hydroxyl group. Typical acyl radicals are acetyl, CH_3CO—, benzyl, C_6H_5CO—, etc.

ACYLATION. A reaction or process whereby an acyl radical, such as acetyl, benzoyl, etc., is introduced into an organic compound. Reagents often used for acylation are the acid anhydride, acid chloride, or the acid of the particular acyl radical to be introduced into the compound.

ADAMANTINE COMPOUND. A compound having in its crystal structure an arrangement of atoms essentially that of diamond, in which every atom is linked to its four neighbors mainly by covalent bonds. An example is zinc sulfide, but it is to be noted that the eight electrons involved in forming the four bonds are not provided equally by the zinc and sulfur atoms, the sulfur yielding its six valence electrons, and the zinc, two. This is the structure of typical semiconductors, e.g., silicon and germanium.

ADAPTATION (Ecology). The process of modification of the living organism to adjust it to the conditions of its environment. Also, an inherited character that enables the organism to meet certain environmental conditions.

All living things are adapted for a mode of life characteristic of their kind, under equally characteristic environmental conditions. They receive from previous generations a heritage that fits them for this mode of life, and all characters in the hereditary complex that are of definite use are adaptive. Wings, for example, are essential flight adaptation, and fins or other similar appendages are commonly found as adaptations for swimming.

Regardless of its adaptive heritage, however, each individual encounters some fluctuations in its environment to which it must adjust itself. The resulting changes in its body are adaptive, no less than its inherited structures. They are the acquired characters of biological literature, and have also been called individual adaptations. Human beings commonly experience two fine examples of this kind of adaptation in the calluses formed by the skin in response to friction, and the deposition of pigment, or tanning, as a protection against excessive ultraviolet light. A less evident result of exposure to ultraviolet light is a protective thickening of the epidermis, probably as important as the accompanying increase in pigmentation. It is interesting to note that the dark-skinned natives of central Africa, for example, do not sunburn as easily as persons with light-colored skin.

Although of different species, it is interesting to note that animals and plants from different parts of the earth often appear to be related as the result of certain adaptations to a given regional environment. The resemblance of cacti which are prevalent in the deserts of the New World, but unknown in the deserts of Africa, nevertheless look very much like the euphorbias found in Africa. Botanically they are unalike. The cava found in the Argentine pampas is related to the guinea pig, although its resemblance to the jack rabbit of similar grassland environment in North America is indeed striking. Adaptation of this type is referred to as *convergent adaptation*.

These adptations have long been recognized, bringing about a series of generalizations which usually apply with few exceptions. For example, Gloger's rule states that cold dry climates encourage light colorations in animals whereas warm and moist climates encourage darker colors. Allen's rule states that parts protruding from the body tend to be shorter in colder climates (long noses are not found among the natives of cold central Asia, for example). Bergmann's rule states that individuals tend to be smaller in warmer climates.

ADAPTIVE CONTROL. A closed-loop control system in which system performance is monitored in relation to what might be termed an index of performance. Adaptive control is a sophisticated improvement, in most cases, of what might be called conventional control systems which employ feedback and sometimes feedforward concepts. Prior to considering adaptive control, some readers may wish to turn to the entry on **Control System (Automatic)** for a fundamental review of the principles of automatic control. Reduced to basics, adaptive control is, in essence, a superimposition of an additional measurement and decision-making system for "controlling the controller." Thus, adaptive systems are sometimes called "self-tuning" controllers. An example is in order to illustrate why this additional control system complication is sometimes needed.

Consider a heat exchanger that uses saturated steam to heat water that flows through its tube bundle. A simple controller will sense the outlet water temperature and attempt to position the steam valve so that the actual water temperature will equal the desired water temperature. This system, however, will not accommodate the nonlinearities in the steam valve and possible changes in stream pressure. A *cascade controller* is a partial approach to adaptive control and is described in entry on **Cascade Control.** In our heat exchanger example, a cascade control system will use a slower-acting temperature controller, not to directly control the steam supply valve, but rather to adjust the set point of a faster-acting steam controller. In this example, one controller is superimposed over another controller.

Unfortunately, a fixed-parameter temperature controller encounters difficulties because of the nonlinear, time-varying behavior of the process. For example, a change in water flow rate changes the effective delay time and heat transfer characteristics of the process. Gradual fouling of the heat exchanger tubes also changes the process dynamics over time. As a result, good control performance at one operating condition can give way to very poor performance (overdamped or unstable response) at another operating condition.

In adaptive control, a method must be provided for the system to change parameters automatically within the closed loop. Auxiliary control variables are measured and evaluated with the objective of modifying the principal control functions so that selected control system performance criteria or indexes can be better realized. Adaptive control is similar to optimal control, in that various parameters are changed as a function of time. Adaptive control differs from optimal control in that parameters in the model of the process are to be evaluated *on-line*. Thus, adaptive control *combines control with the solution* of the identification problem. The resulting control may or may not be optimal. In particular, adaptive control is useful in cases where the process dynamics are not fully defined or change with time.

Professional thinking in terms of adaptive control of processes and machines has been revised, particularly during the late 1980s as the result of introduction of microprocessors and minicomputers to the hardware of automatic control. Adaptive control is part of a family of concepts that fall under the general umbrella term of *expert systems* or *smart controllers*. Some of the rudimentary concepts of artificial intelligence (AI) may be involved. See entry on **Artificial Intelligence.**

Fundamental to the consideration of adaptive control are: (1) sufficient betterment in control performance must be obtainable to justify the additional computation required, often extensive, (2) the state variables or the model parameters that are to be evaluated must be observable directly or indirectly, (3) where parameters are not directly computable from measured signals, there must be sufficient variation on inputs to permit the estimation of these parameters by their influence, and (4) the dynamics of the system must be such that the parameters can be evaluated in a time that is reasonable based upon their rate of change. The fundamental operation of adaptive control is shown in Fig. 1.

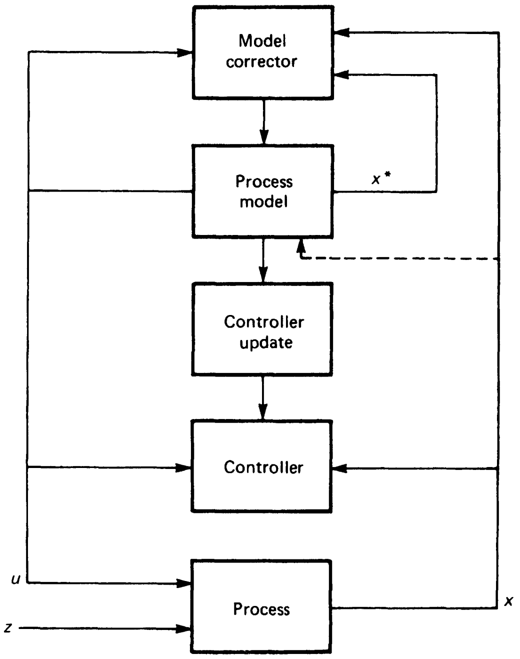


Fig. 1. Hierarchical control structure (x, state factor, x^*, new state factor, u, control factor; z, z-transform)

The basic concepts can be demonstrated by use of a model-adaptive system, such as shown in Fig. 2. In the model-adaptive system, the controlled variable c is compared with the model output c_d. The difference in the two signals is a measure of performance or performance error. Control parameters are adjusted as a function of the performance error to bring the controlled variable to the desired value. One advantage of this type of system is the independent operation of the adaptive and main control loops. Operation of the main control loop can be maintained in event of a failure in the adaptive loop. The complexity of adaptive systems is a major disadvantage in many practical applications. Because of the feedback, which is a part of all closed-loop systems, there is a problem of stability against oscillation. Various forms of Nyquist criteria may be applied in the design of closed-loop systems to ensure satisfactory performance from the standpoint of stability.

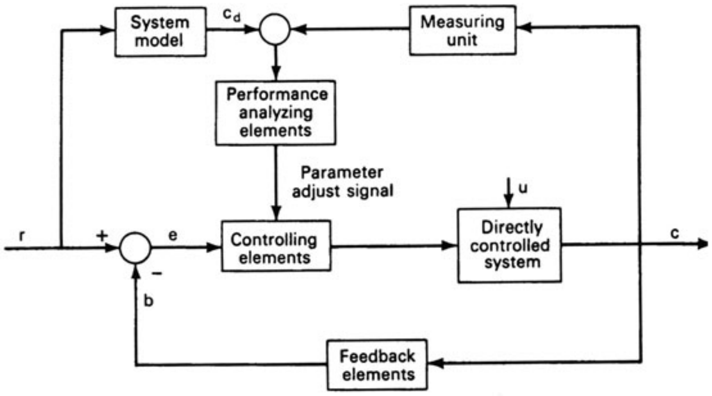


Fig. 2. Model adaptive system.

The adaptive concept as applied to a 2-axis milling machine is shown in Fig. 3. The cutter velocity signal ω and torque signal τ are used in a simple multiplier to obtain a signal proportional to the power delivered to the workpiece by the cutter. From a previously determined relationship between cutter power and workpiece feed rate (axis velocity), a model signal that is useful in determining maximum cutter deflection is obtained. This signal is used to modify position-loop command (axis velocity) for optimum conditions of stock removal. The adaptive control loop for a milling machine is shown in Fig. 4.

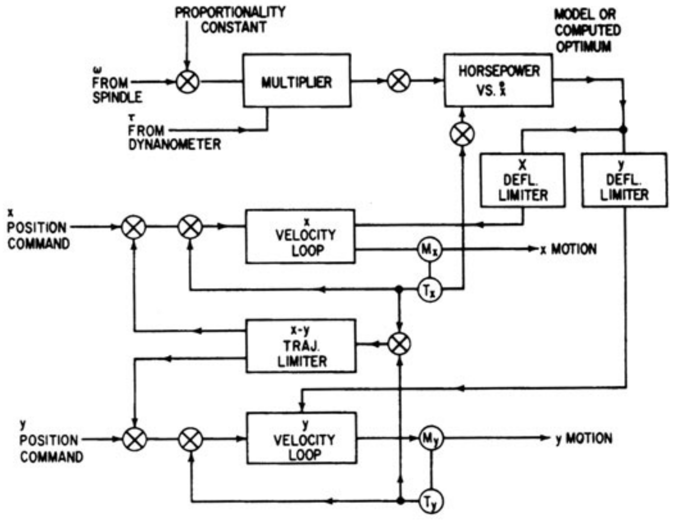


Fig. 4. Adaptive control system for milling machine.

Adaptive control as applied to the heat balance problem of a semibatch reactor is shown in Fig. 5. Reactant gas A is added to liquid B to make liquid product C, also with the production of liquid by-product D. For maximum yield of C, it is desirable to operate the reactor near the batch freeze point. The setpoint of the temperature control loop must be adjusted to follow the freeze point. The freeze point increases with changing chemical composition during the batch. As shown, an analog computer is used to compute the incoming flow and thus determines the freeze point of the batch system to be used as the setpoint of the temperature controller. A different approach to the problem is shown in Fig. 6. This system does not depend upon external data. Rather, the adaptive control signal which controls the primary loop gain is obtained from analysis of the error signal. The signal is separated into high and low components, multiplied by a ratio factor, combined, and then passed through a filtering and integrating network. A system of this type may be designed for universal purposes and thus the performance characteristics and reliability may not be expected to compete with

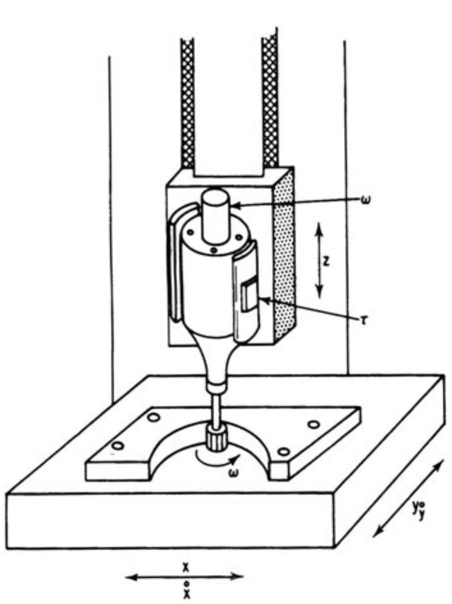


Fig. 3. Variables involved in adaptive control system for a 2-axis milling machine.

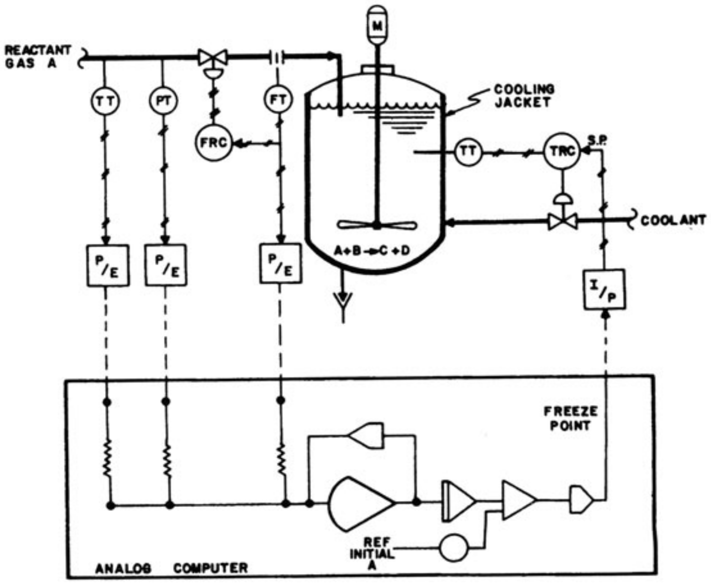


Fig. 5. An adaptive control approach to heat-balance problem of a semibatch reactor.

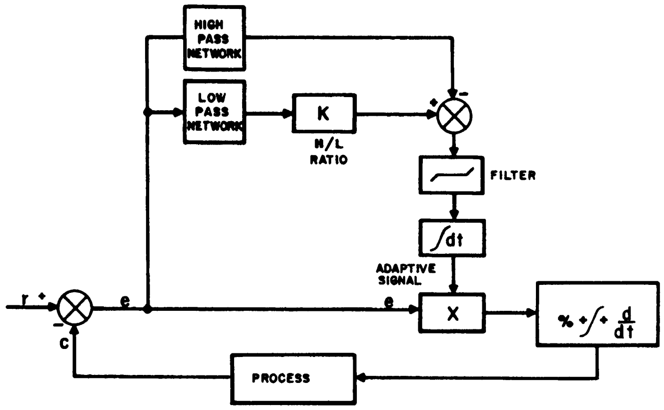


Fig. 6. Adaptive control approach to heat-balance problem based upon analysis of error signal.

the programmed controller, inasmuch as the latter is specifically designed for the particular process.

ADDAX (*Mammalia, Artiodactyla*). A screw-horned antelope, *Addax nasomaculatus*, of northern Africa and Arabia.

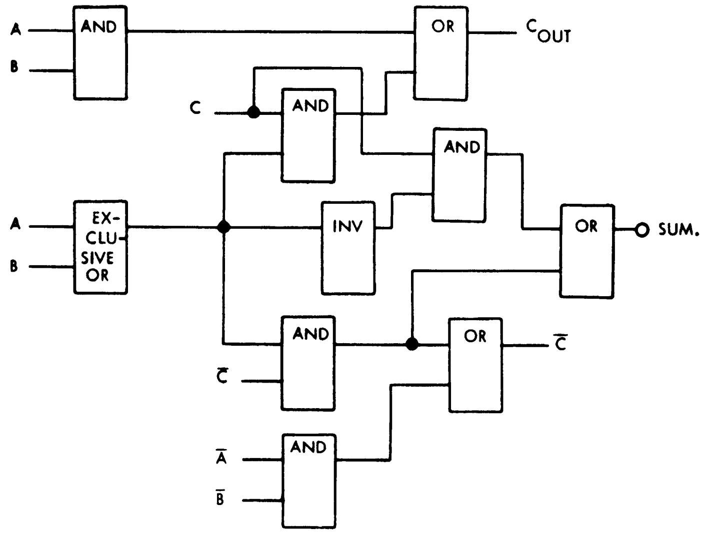
Built like a reindeer, these animals have a short neck, long rump, medium-long sturdy legs, and a straight back, with the withers slightly higher. The rather large head has a considerable tuft on the forehead, especially in old males. The eyes are small and the preorbital glands are not visible. The moderately thin horns, which slant backward and up and slightly to the sides in large, flat spirals are almost round in a crosscut. They have low, flat hooves with flat plantars (sole of foot) which are not sectioned into balls. The round, thin tail has a short tassel on the end. The winter coat is longer and rougher than the summer coat.

Like the scimitar-horned oryx and the North African hartebeest, the addax was kept semi-tame in large numbers in ancient Egypt. According to pictures from this time, addax stood in stables like goats. They were fed out of troughs, led on a bridle, and were probably slaughtered for cult purposes. Keepers developed special procedures for protection against the pointed horns of the animal.

Since 1900 the addax has disappeared from Egypt, and even before that time it had been exterminated in the north of Algeria, Tunisia, Libya, and surrounding environs. Formerly, there were herds of hundreds and thousands. They may go for weeks and even months without water. Like camels, the walls of their stomachs are said to be transformed into alveolar reservoirs for liquids. The animals are active in the morning, evening, and night. In order to protect themselves against strong winds and solar radiation, they dig holes with their front legs and lie down in them.

Although in the past, addax antelopes were commonly kept in zoos, this is now rare. Thus, the breeding groups in a comparatively few zoos are important to the ultimate preservation of the species. In captivity, the addax achieves a maximum age of about 18 years.

ADDER (Computer System). A digital circuit which provides the sum of two or more input numbers as an output. A one-bit binary adder is illustrated in the accompanying figure. In this diagram, *A* and *B* are the input bits and *C* and $\overline{C}$ are the carry and no carry bits from the previous position. There are both serial and parallel adders. In a serial adder, only one adder position is required and the bits to be added are sequentially gated to the input. See also **Gate (Computer System).** The carry or no carry from the prior position is remembered and provided as an input along with the bits from the next position. In a parallel adder, all the bits are added simultaneously with the carry or no carry from the lower-order position propagated to the higher position. In a parallel adder, there may be a delay due to the carry propagation time. See also **Half-Adder.**



Binary adder.

An adder may perform subtraction as well as the addition of two numbers. Generally, this is effected by complementing one of the numbers and then adding the two factors. The following is an example of a two's-complement binary-subtraction operation.

(a)		0 1 1 0	+6	(True)
	(+)	1 0 1 0	−6	(Complement)
		1 0 0 0 0	0	(True)
(b)		0 1 0 1	+5	(True)
	(+)	1 0 1 0	−6	(Complement)
		1 1 1 1	−1	(Complement)
(c)		1 1 1 1	(Complement) = −0 0 0 1 (True)	

The two's complement of a binary number is obtained by replacing all **1**'s with **0**'s, all **0**'s with **1**'s, and adding **1** to the units position. In (a) above, **6** is subtracted from **6**, and the result is all **0**'s; the carry implies that the answer is in true form. In (b), **6** is subtracted from **5**, and the result is all **1**'s with no carry. The no carry indicates the result is in complement form and that the result must be recomplemented as shown in (c).

Thomas J. Harrison, International Business Machines Corporation, Boca Raton, Florida.

ADDISON'S DISEASE. A disease caused by malfunction of the adrenal glands. It is characterized by a bronze color of the skin, prostration, anemia, disturbance of electrolyte metabolism, and diarrhea. Many cases are believed to result from an autoimmune disorder, i.e., an immune reaction by the body against some stimulus naturally present within itself. Sometimes it is caused by involvement of the glands by tuberculosis and other infections, metastatic tumors, or amyloidosis, but the principal cause is primary adrenal failure.

Success in treating the disease in recent years has increased because of the availability of cortisol or cortisone and also because of supplementation with a sodium-retaining hormone to maintain the salt level in the body. The availability of purified hormones has simplified the treatment, where often each symptom can be alleviated with a corresponding hormone. For example, there is a specific hormone responsible for maintaining the water and salt balance of the body. Past practice has also included taking meals high in starch at fairly frequent intervals.

Recent studies by Sadeghi-Nejad and Senior have shown that, in children, *adrenoleukodystrophy* is an additional, genetically determined, rare cause of Addison's disease. One form of adrenoleukodystrophy in infants is transmitted as an autosomal recessive trait; the disorder in older patients is sex-linked. Adrenoleukodystrophy is a peroxisomal dysfunction, resulting in the impaired oxidation of very-long-chain saturated fatty acids, mainly hexacosanoic acid. These fatty acids accumulate in the brain, adrenal glands, and other organs and are believed to be responsible for the clinical manifestations of the disease.

The neurologic features of adrenoleukodystrophy include dementia and deterioration of vision, speech, and gait.

Additional Reading

Edwards, C. R. W.: "Addison's Disease" in *Clinical Endocrinology*, Besser and Cudworth, Editors, Lippincott, Philadelphia, Pennsylvania, 1987.

Sadeghi-Nejad, A., and B. Senior: "Adrenomyeloneuropathy Presenting in Addison's Disease in Childhood," *N. Eng. J. Med.*, 13 (January 4, 1990).

Schutgens, R. B., et al.: "Peroxisomal Disorders: A Newly Recognized Group of Genetic Diseases," *Eur. J. Pediatrics*, *133*, 430 (1986).

ADDITION. A fundamental operation for combining mathematical terms. The symbol + is used to denote the operation. To perform the operation of addition of two quantities x and y, it is necessary to identify like terms. If x and y are polynomials in z, then like terms are the terms in x and y containing z to the same power. For complex numbers, $x = a + ib$, $y = c + id$, $x + y = (a + c) + i(b + d)$. For fractions to have like terms, one must use common denominators.

$$\frac{a}{b}+\frac{c}{d}=\frac{ad}{bd}+\frac{cb}{bd}=\frac{ad+cb}{bd}.$$

Frequently, common factors can be cancelled in the numerator and denominator, e.g.,

$$\frac{1}{3}+\frac{1}{6}=\frac{6+3}{18}=\frac{9}{18}=\frac{1}{2}$$

Addition consists of finding the algebraic sum of like terms. Eventually, one must use the same process as in arithmetic: determine the numerical value of an addition by the addition tables for the base in use.

Donald R. Hodge, Alexandria, Virginia.

ADDITIVE COLOR PROCESS. An early system of color imagery in which the color synthesis is obtained by the addition of colors one to another in the form of light rather than as colorants. This color addition may take place (1) by the simultaneous projection of two or more (usually three) color images onto a screen, (2) by the projection of the color images in rapid succession onto a screen or (3) by viewing minutely divided juxtaposed color images.

In the case of a three-color process, three color records are made from the subject recording, in terms of silver densities, the relative amounts of red, green and blue present in various areas of the subject.

When the additive synthesis is to be made by simultaneous projection, positives are made from the color separation negatives and projected with a triple lantern onto a screen through red, green and blue filters. The registered color images give all colors of the subject due to simple color addition, red plus green making yellow, red plus blue appearing magenta, etc.

When the additive synthesis is made by successive viewing, the same three color images must be flashed onto the screen in such rapid succession that the individual red, green and blue images are not apparent. Simple color addition is again obtained but this time use is made of the persistence of vision to "mix" the colors. See accompanying diagram.

Mechanism of color addition.

The third type of additive synthesis makes use of the fact that small dots of different colors, when viewed from such a distance that they are no longer individually visible, form a single color by simple color addition. The three color images in this type of process are generally side-by-side in the space normally occupied by a single image. The red record image will be composed of a number of red dots or markings of differing density which in their sum total will compose the red record image. Alongside the red markings will be green and blue markings, without any overlapping. When viewed at such a distance that the colored markings are at, or below, the limit of visual resolution, the color sensation from any given area will be the integrated color of the markings comprising the area—an additive color mixture.

ADDITIVE (Fuel). See **Petroleum.**

ADDITIVES (Food). The use of food additives is centuries old. One of the earliest and still most widely practiced example is the addition of common salt to foods, either by a food preparer or processor, or by the consumer at the table. Normally, the food producing professionals and the lay public do not view salt as a food additive, probably because salt is regarded as a substance widely found in nature, and particularly in seawater. Nevertheless within the last 10 to 20 years, there has been a growing awareness that salt is truly an additive and to certain controlled diets can have an adverse role as well as the numerous positive roles that it plays in food processing and preservation.

The term *food additive* is most frequently associated with the more complex substances, the presence or function of which the average consumer seldom notes unless the package labels are read. Particularly where additives contribute to the pleasure of the consumer, they are commonly passed over unnoticed. By far, the majority of consumers prefer breads that do not stale readily, salad dressings that do not separate into layers after a period of shelf life, dairy products that do not become rancid after a few days, and meats that retain their color and wholesome appearance and taste after relatively long periods of refrigeration. These food qualities, of course, are made possible because of the many advances over the years in food additive technology.

Additives also contribute to the ease of processing and marketing foods and, even though indirectly, have played a positive role in terms of food transportation, distribution, and inventorying costs. Conversely, some additives in the past have been found to be damaging to health. Some of the colorants formerly used, for example, have been banned because of their carcinogenic qualities when tested in laboratory animals. Considering the vast quantities of additives used, the majority used to date have not been shown to be harmful.

Some government regulators, as in the United States, have established a class of ingredients known as GRAS (Generally Recognized as Safe) or by some similar designation. See **GRAS.**

There are tens of categories of food chemicals, most of which are described in various entries throughout this encyclopedia. Consult the following topics in the alphabetical index: Acidulants, alkalizers, anticaking agents, antimicrobial agents, antioxidants, bleaching agents, bodying and bulking agents, buffers, carriers, clarifying agents, coating agents, colorants (Annato food colors; anthocyanins; betalaines; carotenoids), enzyme preparations, firming agents, flavorings and flavor enhancers and potentiators, humectants and moisture-retaining agents, leavening agents, masticatory substances, oxidizing agents, pectins, polymeric food additives, and sweeteners.

Dispersing, emulsifying, gelling, thickening, and stabilizing agents are covered under **Colloid System.** Dough conditioners are described under **Oxidation and Oxidizing Agents.** Nonnutritive sweeteners are included in entry on **Sweeteners.**

Additional Reading

Carroll, L. E.: "Stabilizer Systems Reduce Texture Problems in Multicomponent Foods and Bakery Products," *Food Technology*, 94 (April 1990).

Dziezak, J. D.: "Acidulants (Foods)," *Food Technology*, 75 (January 1990).

Dziezak, J. D.: "Phosphates Improve Many Foods," Food Technology, 79 (April 1990).

Glicksman, M.: "Hydrocolloids and the Search for the 'Oily' Grail," *Food Technology*, 94 (October 1991).

Irwin, W. E.: "Isomalt—A Sweet, Reduced-Calorie Bulking Agent," *Food Technology*, 128 (June 1990).

Kokini, J. L., Lai, L. S., and L. L. Chedid: "Effect of Starch Structure on Starch Rheological Properties," *Food Technology*, 124 (June 1992).

Oser, B. L., and R. A. Ford: "FEMA (Flavor and Extract Manuturers Association) Expert Panel: 30 Years of Safety Evaluation for the Flavor Industry," *Food Technology*, 84 (November 1991).

Pennington, N. L., and C. W. Baker: "Sugar; A User's Guide to Sucrose," Van Nostrand Reinhold, New York, 1990.

Pszczola, D. E.: "Oat-Bran-Based Ingredient Blend Replaces Fat in Ground Beef and Pork Sausage," *Food Technology*, 60 (November 1991).

Staff: "Powdered Silicone Antifoam Reduces Unwanted Foam in Food Applications," *Food Technology*, 132 (January 1991).

Summerkamp, B. S., and M. Hesser: "Fat Sutstitute Update," *Food Technology*, 92 (March 1990).

Tye, R. J.: "Konjac Flour: Properties and Applications," *Food Technology*, 82 (March 1991).

Waniska, R. D., and M. H. Gomez: "Dispersion Behavior of Starch," *Food Technology*, 110 (June 1992).

Yackel, W. C., and C. Cox: "Application of Starch-Based Fat Replacers," *Food Technology*, 146 (June 1992).

ADDRESS (Computer System). An identification, representated by a name, label, or number, for a digital computer register, device, or location in storage. Addresses are also a part of an instruction word along with commands, tags, and other symbols. The part of an instruction which specifies an operand for the instruction may be an address.

Absolute address or *specific address* indicates the exact physical storage location where the referenced operand is to be found or stored in the actual machine code address numbering system.

Direct address or *first-level address* indicates the location where the referenced operand is to be found or stored with no reference to an index register.

Indirect address or *second-level address* in a computer instruction indicates a location where the address of the referenced operand is to be found. In some computers, the machine address indicated can in itself be indirect. Such multiple levels of addressing are terminated either by prior control or by a termination symbol.

Machine address is an absolute, direct, unindexed address expressed as such, or resulting after indexing and other processing have been completed.

Symbolic address is a label, alphabetic or alphameric, used to specify a storage location in the context of a particular program. Sometimes programs may be written using symbolic addresses in some convenient code, which then are translated into absolute addresses by an assembly program.

Base address permits derivation of an absolute address from a relative address.

Effective address is derived from applying specific indexing or indirect addressing rules to a specified address.

Four-plus-one address incorporates four operand addresses and a control address.

Immediate address incorporates the value of the operand in the address portion instead of the address of the operand.

N-level address is a multilevel address in which N levels of addressing are specified.

One-level address directly indicates the location of an instruction.

One-plus-one address contains two address portions. One address may indicate the operand required in the operation. The other may indicate the following instruction to be executed.

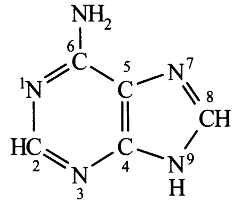
Relative address is the numerical difference between a desired address and a known reference address.

Three-plus-one address incorporates an operation code, three operand address parts, and a control address.

Zero-level address enables immediate use of the operand.

Thomas J. Harrison, International Business Machines
Corporation, Boca Raton, Florida.

ADENINE. A prominent member of the family of naturally occurring purines with the accompanying formula. Adenine occurs not only in ribonucleic acids (RNA), and deoxyribonucleic acids (DNA), but in nucleosides, such as adenosine, and nucleotides, such as adenylic acid, which may be linked with enzymatic functions quite apart from nucleic acids. Adenine, in the form of its ribonucleotide, is produced in mammals and fowls endogenously from smaller molecules and no nutritional essentiality is ascribed to it. In the nucleosides, nucleotides, and nucleic acids, the attachment or the sugar moiety is at position 9.

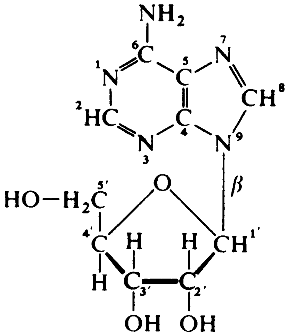


Structure of adenine.

The purines and pyrimidines absorb ultraviolet light readily, with absorption peaks at characteristic frequencies. This has aided in their identification and quantitative determination.

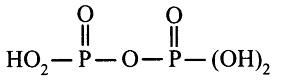
ADENOMA. A benign tumor consisting of an encapsulated overgrowth of epithelial cells of a glandular structure. Adenomata may occur in the endocrine glands, the gastro-intestinal tract, the respiratory system, the breast, and wherever glandular epithelium occurs. A malignant adenoma is referred to as an *adenocarcinoma*.

ADENOSINE. An important nucleoside. The upper portion of the accompanying formula represents the adenine moiety, and the lower portion of the pentose, D-ribose.



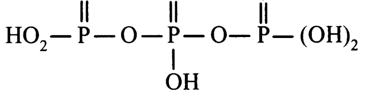
Structure of adenosine.

ADENOSINE PHOSPHATES. The adenosine phosphates include *adenylic acid* (adenosine monophosphate, AMP) in which adenosine is esterified with phosphoric acid at the 5′-position; *adenosine diphosphate* (ADP) in which esterification at the same position is with pyrophosphoric acid,



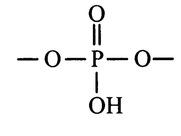
$$HO_2-\overset{\overset{\displaystyle O}{\|}}{P}-O-\overset{\overset{\displaystyle O}{\|}}{P}-(OH)_2$$

and *adenosine triphosphate* (ATP) in which three phosphate residues



$$HO_2-\overset{\overset{\displaystyle O}{\|}}{P}-O-\underset{\underset{\displaystyle OH}{|}}{\overset{\overset{\displaystyle O}{\|}}{P}}-O-\overset{\overset{\displaystyle O}{\|}}{P}-(OH)_2$$

are attached at the 5′-position. Adenosine-3′-phosphate is an isomer of adenylic acid, and adenosine-2′,3′-phosphate is esterified in two positions with the same molecules of phosphoric acid and contains the radical



$$-O-\underset{\underset{\displaystyle OH}{|}}{\overset{\overset{\displaystyle O}{\|}}{P}}-O-$$

ADENOSIS. Any disease of a gland or glands.

ADHARA (ϵ Canis Majoris). Ranking twenty-second in apparent brightness among the stars, Adhara has a true brightness value of 8,000 as compared with unity for the sun. Adhara is a blue-white, spectral type B star and is located in the constellation Canis Majoris. Estimated distance from the earth is 600 light years. See also **Constellations.**

ADHESION (Physics). The terms adhesion and cohesion designate intermolecular forces holding matter together. The tendency of matter to hold itself together or to cling to other matter is one of its most characteristic properties. Adhesion and cohesion are merely different aspects of the same phenomenon, which is apparently of the nature of an intermolecular attraction. One speaks of cohesion as an interaction between adjacent parts of the same body and as acting throughout the interior of its substance, while adhesion refers to a similar interaction between the closely contiguous surfaces of adjacent bodies.

In the case of solids, the experimental study of adhesion and cohesion, in which two solids are brought into intimate contact, can depend on the nature of adsorbed gases. For clean surfaces it is necessary to carry out the experiments in "ultra high vacuum" at pressures better than 10^{-10} torr.

There is reason to believe that as two neutral molecules or atoms approach each other, their mutual potential energy reaches a minimum value at a certain equilibrium distance; so that work would be necessary either to push them closer or to pull them farther apart, because of forces which are probably electrical. See **Least Energy Principle.** The distribution of molecules, ions, or atoms in a solid is determined by this type of equilibrium, and the regular spacing of crystal structure and the architecture of the molecule itself are dependent upon it. Any force tending to diminish the equilibrium distance meets with the rapidly increasing reaction of compressive elasticity, while any force tending to increase it is opposed by cohesion, which increases at first and then rapidly diminishes toward zero as the point of fracture is reached.

The behavior of bodies which are aggregates of crystals or of fibers is complicated by the friction and the adhesion of the adjacent particles, so that the ultimate strength of a material is not a safe measure of its true cohesion. A filament of spun quartz may be much stronger when freshly drawn than later when crystallization replaces its initial cohesion by the adhesion between separate crystals; and yarn is not nearly so strong as the cotton or wool fiber composing it.

Adhesion increases with closenss of contact. This explains why one must bear down with a pencil to make a mark on paper, why fine dust adheres more firmly than coarse sand, and why a liquid or a gum usually sticks to a solid better than another solid does.

Cohesion in liquids is usually less, and in gases it is always much less, than in solids. Aside from the pressure in liquids due to external causes, there is presumably a very great internal or intrinsic pressure, due to intermolecular attraction, but not capable of direct measurement by means at our disposal. The clearest evidences of its existence are the work required for thermal expansion and the phenomenon of surface tension.

Very practical aspects of materials adhesion are encountered in the bulk materials processing industries. Two common terms are used—*arching* and *flushing*. Both properties are associated with what is known as the *angle of repose*. See also **Repose (Angle of)**. Many materials, such as lampblack, activated carbon, zinc oxide, titanium oxide, fine soda ash, and hydrated lime, among many others, tend to arch when placed in containing vessels such as hoppers. Materials above the arch are held back and consequently interfere with the unloading of the hopper and thus prevent uniformity of solids-flow, as required by most continuous processes. Arching is best overcome by keeping the individual solid particles constantly in motion. Electric vibrators are commonly used for this purpose. Very few materials are crushed or pulverized so well that all particles are of approximately equal size. In the absence of hopper vibration, the smaller particles and fines tend to segregate and roughly collect in a center cone, while the larger particles roll to the side. Thus, the fines drop out first, while the larger particles discharge later, the end result of which is formation of an arch. Flushing is caused by the sudden breaking of an arch or otherwise clogged state. Thus, vibration also helps to eliminate flushing.

In a very interesting article, Jearl Walker (*Sci. Amer.*, January 1982, pp. 174–179) explains in some detail the differences between sand castles and mudpies. The author explains why sand castles disintegrate when they dry and mud pies do not! Also, why mud pies shrink when they dry, and sand castles do not!

ADHESION (Work of). The work of adhesion W_{AB} between two liquids A and B is the increase in free surface energy on separating 1 cm^2 of interface AB

$$W_{AB} = \gamma_A + \gamma_B - \gamma_{AB}$$

where γ_A and γ_B are the surface tensions of A and B respectively against their vapors, and γ_{AB} is the interfacial tension. For a solid-liquid interface the work of adhesion W_{SL} is defined as the work required to separate 1 cm^2 of interface in a vacuum to give a naked solid surface

$$W_{SL} = \gamma_S + \gamma_L - \gamma_{SL}$$

where γ_S and γ_L are the surface tensions measured in a vacuum.

It may be shown that

$$W_{SL} = \gamma_S - \gamma_{SV0} + \gamma_L (1 + \cos \theta_E)$$

where γ_{SV_0} is the surface tension of the solid covered by an absorbed film of liquid in equilibrium with the vapor, and θ_E is the equilibrium contact angle.

ADHESIVES. Materials or compositions which enable two surfaces to become united are known as adhesives. Years ago the terms *adhesive* and *glue* were considered synonymous, but the general term glue now implies a sticky substance, whereas many adhesives are not sticky. Adhesives may be broadly classified into two main groups; organic and inorganic adhesives with the organic materials being subdivided into those of animal origin, vegetable origin, and synthetic origin. A more useful classification is one based on the chemical nature of the adhesive. Such a chemical classification comprises (1) protein or protein derivatives, (2) starch, cellulose, or gums and their derivatives, (3) thermoplastic synthetic resins, (4) thermosetting synthetic resins, (5) natural resins and bitumens, (6) natural and synthetic rubbers, and (7) inorganic adhesives.

See also **Epoxy Resins**.

Adhesives Advancements. During the early 1990s, adhesives are continuing to play an ever-increasing role as an engineering material for joining parts and pieces, even in industries where, traditionally, joining processes have been essentially synonymous with welding, soldering, brazing, and other metal-joining techniques. The greater use of lower-gauge sheet metal has made it increasingly difficult to control spot welding, and this has forced engineers to turn to other joining techniques. The engineering of composite materials has been made possible in thousands of instances because of the availability of a wide variety of adhesive materials. The trend toward the greater use of adhesives also has been prompted by the continued replacement of metal parts by plastic materials, as exemplified by the automotive and appliance manufacturers.

Thermoplastic hot-melt adhesive has grown at a pace exceeding that of other adhesive categories, mainly because of the simplicity of the fast heat-and-stick process, with no need to mix multiple components in critical ratios. However, for some applications they have been limited by low cohesion (strength within the adhesive itself) and by temperature limitations of gun-applied material (maximum, 230°F; 110°C). These limitations have been overcome by the reactive hot-melts. These newer materials have a number of advantages: (1) higher-temperature performance (up to 400°F; 204°C); (2) improved chemical resistance, including the ability to withstand numerous solvents, plasticizers, and boiling water; (3) longer open time (1 to 3 minutes) for positioning parts during application; and (4) lower application temperatures (200° to 300°F; 93° to 149°C), as compared with earlier hot-melt adhesives.

Epoxies, commonly used for years in aircraft manufacture for bonding thin aluminum sheet and honeycomb materials, have been subjected to extensive research in recent years. Much effort has been directed toward curing epoxy bonds, including the development of an ultraviolet laser curing process.

Much research has been directed toward improving one-component adhesives. Film adhesives are available that confer electrical conductivity as well as joint strength for integrated circuit chip bonding. Film adhesives now confer electrical conductivity as well as joint strength. Over the past decade, the automotive industry has highlighted application research in the interest of achieving improved, simpler, and lighter-weight bonds in automobile and truck manufacture.

Approximate Classification of Adhesives. Adhesives may be classified according to the purpose for which they are intended or by the constitution of their ingredients.

In terms of uses, there are (a) for bonding rigid surfaces, such as wood, glass, porcelain, rigid plastics, or metal, and (b) for bonding flexible surfaces such as paper, textiles, leather, flexible plastics, thin metallic sheets, and the like. In the latter category it is necessary that the adhesive have as high a degree of flexibility as the less flexible of the bonded surfaces.

The protein and protein derivative adhesives include those made from casein, zein, soybean proteins, and other proteins and the very important groups of glues made from hides, bones, etc., fish glues from fish offal, and those from blood albumen.

The group of adhesives made from materials such as starch and the vegetable gums are also known as vegetable adhesives or glues. They comprise the adhesives made from starch and processed starch, the dextrins, and the water soluble gums such as gum arabic, ghatti, tragacanth, Indian gum and the like. These are water soluble or water suspendable materials. The starch and dextrin products are used for the bonding of paper, wood, and textiles; the gum products are used as adhesives for paper, postage stamps and other stamps, and for miscellaneous items.

Cellulose adhesives are principally cellulose derivatives such as methylcellulose, ethylcellulose, cellulose acetate and nitrate, and sodium carboxymethyl cellulose. These products are used as components in adhesive compositions such as those used for the bonding of leather, cloth, paper, and many other materials.

Thermoplastic synthetic resin adhesives comprise a variety of polymerized materials such as polyvinyl acetate, polyvinyl butyral, polyvinyl alcohol, and other polyvinyl resins; polystyrene resins; acrylic and methacrylic acid ester resins; cyanoacrylates; and various other synthetic resins such as polyisobutylene, polyamides, coumarone-indene products, and silicones. Such thermoplastic resins usually have permanent solubility and fusibility so that they creep under stress and soften when heated. They are used for the manufacture of tapes, safety glass, and shoe cements and for the bonding of foils, metals, woods, rubber, paper, and many other materials.

Thermosetting synthetic resin adhesives comprise a variety of phenolaldehyde, urea-aldehyde, melamine-aldehyde, and other condensation-polymerization materials like the furane and polyurethane resins. Thermosetting adhesives are characterized by being converted to insoluble and infusible materials by means of either heat or catalytic action. Adhesive compositions containing phenol-, resorcinol-, urea-, melamine-formaldehyde, phenolfurfuraldehyde, and the like are used for the bonding of wood, textiles, paper, plastics, rubbers, and many other materials.

The adhesives of the natural resin and bitumen group consist of those made from asphalts, shellac, rosin and its esters, and similar materials. They are used for the bonding of various materials including minerals, linoleum, and the like.

There are a number of different types of rubber adhesives. Some are simply solutions of rubber latex, rubber, or synthetic and modified rubbers in a solvent, and others are compositions consisting of one of the aforementioned and casein or a synthetic resin. Such compositions are very widely used from the bonding of flexible materials like paper, textiles, leather and rubber to rigid materials like metals and plastics.

The inorganic adhesives are a very important group. They comprise: sodium silicate principally used for corrugated paper and other paper products, laminating metal foils, plywood bonding, and in the production of building and insulating boards; plaster of paris for ceramic and similar products; magnesium oxychloride for ceramics; litharge-glycerin for joints in plumbing fixtures; and Portland cement for bonding mineral aggregates. See also **Silicates (Soluble).**

While it has been convenient to put the various types of materials from which adhesives are made into groups, many adhesive compositions are made of substances or materials from a number of groups.

In this connection it should be noted that the term *cement* is often used as a synonym for adhesive. A cement is a particular kind of adhesive which consists of any material that can be prepared in a plastic form which hardens to bond together various solid surfaces.

An excellent review of "Adhesive Bonding" is contained in the *Modern Plastics Encyclopedia,* issued annually by Modern Plastics, Pittsfield, Massachusetts.

ADIABATIC. Occurring without gain or loss of heat by the system involved: for example, adiabatic expansion. A completely adiabatic process is unrealizable, but it can be closely approximated in practice by providing good thermal insulation and by carrying out the process rapidly so that little time is allowed for heat flow in or out of the system. If a thermally isolated system moves through a series of equilibrium states, i.e., undergoes an adiabatic and reversible process, the locus of the points representing these states on a graph is called an *adiabatic* or *adiabat.* The present tendency is to avoid this usage of adiabatic to mean isentropic. See **Adiabatic Process.**

ADIABATIC PROCESS. Any thermodynamic process, reversible or irreversible, which takes place in a system without the exchange of heat with the surroundings. When the process is also reversible, it is called *isentropic*, because then the entropy of the system remains constant at every step of the process. (In older usage, isentropic processes were called simply adiabatic, or quasistatic adiabatic; the distinction between adiabatic and isentropic processes was not always sharply drawn.)

When a closed system undergoes an adiabatic process without performing work (*unresisted expansion*), its internal energy remains constant whenever the system is allowed to reach thermal equilibrium. Such a process is necessarily irreversible. At each successive state of equilibrium, the entropy of the system S_i, has a higher value than the initial entropy, S_0. Example: When a gas at pressure p_0, temperature T_0, occupying a volume V_0 (see Fig. 1) is allowed to expand progressively into volumes $V_1 = V_0 + \Delta V$, etc., by withdrawing slides 1, 2, etc., one after another, it undergoes such a process if it is enclosed in an adiabatic container. After each withdrawal of a slide, the irreversibility of the process causes the system to depart from equilibrium; equilibrium sets in after a sufficiently long waiting period. At each successive state of equilibrium $U_1 = U_2 = \cdots = U_0$, but $S_0 < S_1 < S_2$, etc.

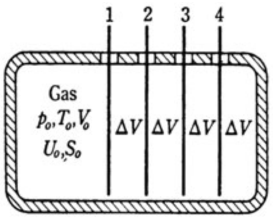


Fig. 1. Successive adiabatic expansions of gas by withdrawing slides.

When an open system in steady flow undergoes an adiabatic process without performing external work, the enthalpy of the system regains its initial value at each equilibrium state, and the entropy increases as before. Example: Successive, *slow* expansions through porous plugs $P_1, P_2 \cdots$ (Fig. 2), when we have

$$H_1 = H_2 = \cdots = H_0$$

but

$$S_0 < S_1 < S_2, \text{ etc.}$$

This process is also necessarily irreversible.

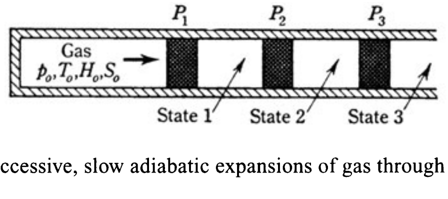


Fig. 2. Successive, slow adiabatic expansions of gas through porous plugs.

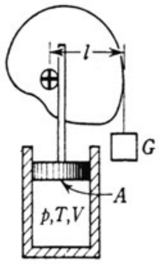


Fig. 3. Isentropic compression (or expansion) in cylinder.

A closed system cannot perform an isentropic process without performing work. Example (Fig. 3): A quantity of gas enclosed by an ideal, frictionless, adiabatic piston in an adiabatic cylinder is maintained at a pressure p by a suitable ideal mechanism, so that $Gl = pA$ (A being the area of piston). When the weight G is increased (or decreased) by an infinitesimal amount dG, the gas will undergo an isentropic compression (or expansion). In this case,

$$S = \text{constant}, \qquad dS = 0$$

at any stage of the process, but

$$U \neq \text{constant}, \qquad H \neq \text{constant}$$

During an isentropic process of a closed system between state 1 and 2, the change in internal energy equals *minus* the work done between the two states, or

$$U_2 - U_1 = -W_{12}$$

work is done "at the expense" of the internal energy.

ADIABATIC WALL. A perfect heat insulator. Since in a rigorous development of the principles of thermodynamics it is necessary to introduce the concept of an adiabatic wall before the concept of heat, it is convenient to adopt the following alternative definition. If two closed systems are place in contact through an adiabatic wall, their states can be varied independently of one another. Any state of one system can co-exist with any state of the other system through such a wall; the systems are not coupled in any way. See **Diathermal Wall.** The number of independent properties of the combined system is equal to the sum of the number of independent properties of the component systems.

ADIPOSE FIN. See **Fishes**

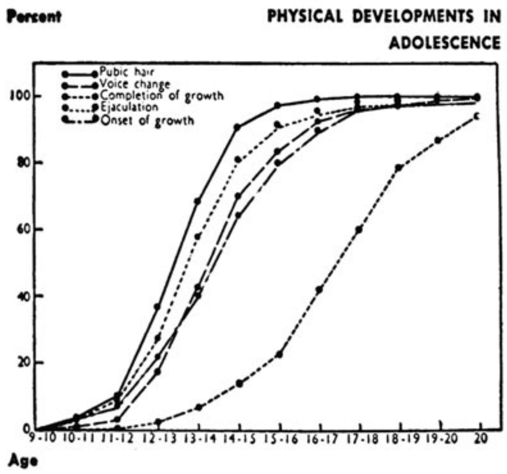
ADOBE. An extremely fine-grained wind-blown clay particularly characteristic of the arid and semi-arid southwestern United States, Mexico, and South America. Used by the southwestern Indians and Mexicans for huts and buildings from prehistoric times.

ADOLESCENCE (Physiology). The period of youth extending from the beginning of puberty to adulthood. For human beings, this period usually ranges between 12 and 20 years. Physical developments in adolescence of the American male are shown in the accompanying chart.

Growth during the first 10 years of life proceeds at a fairly uniform rate. The child's desires usually keep pace with his ability to meet his needs. However, during the adolescent or teenage period, there is disharmony between physiological development, growth, and emotional maturation. The teenager becomes a blend of maturity and immaturity in body, mind, and emotion.

Puberty, which refers to the sexual maturation of the individual, is only part of adolescence. Puberty and adolescence begin in the girl at 11 to 15 years of age. Boys lag about a year behind. The average age for puberty in girls in the United States is around 13.5 years and, in boys, 14.5 years. Adolescence in most American youth terminates at about 19 years of age.

The first half of adolescence is characterized by physical growth and change; the latter half brings more intellectual and emotional changes.



This chart portrays the percent of American males undergoing the indicated changes in physical development at the various ages during adolescence.

Adolescent growth varies widely between individuals. Failure to realize this may cause unnecessary anxiety to both parents and child. Growth rate is influenced by the state of health and nutrition. Not only is there variability in the growth of different individuals, but growth within the single individual is not harmonious. Certain parts of the body grow, while other parts lag behind. The child may grow tall without putting on weight, or he may show more sexual maturity than maturity of the digestive organs. This organ imbalance leads to organ instability. Thus, laziness and awkwardness may result. In the average boy, the period of greatest height increase is from about 12 to 14.5 years, although they continue to grow until about the age of 18. They gain most of their weight between the ages of 13 and 16. Girls grow tallest from about 10.5 to 14 years; maximum weight development is from about ages 11.5 to 14.5. As much as 6 inches (15 centimeters) in height and 25 pounds (11.3 kilograms) in weight may be gained within a single year.

Puberty in the boy is marked by growth of the sexual organs and the beginning of their function. Secondary sexual changes occur. His voice deepens; his shoulders broaden; his muscles harden; his legs lengthen; his hands and feet grow disproportionately large.

Arrival of the first menstrual period in a girl may be shocking and frightening if she is improperly prepared for the event. Menstrual irregularity and pain are common the first year or so, and are usually without significance. The broadening of hips, development of breasts, and appearance of pubic hair usually parallel the beginning of menstruation, but are subject to wide variation. Both sexes are preoccupied with their bodies and unable to take their growth for granted.

Obesity is often a problem in this period and may be a factor in delayed puberty. Whether the condition is the result of poor eating habits, of emotional instability, or of endocrine dysfunction, it merits careful investigation.

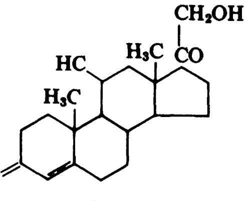
Physical maturity is reached when the body has its final height and has assumed adult proportions. The secondary sexual characteristics are fully developed, and the sexual functions have been established.

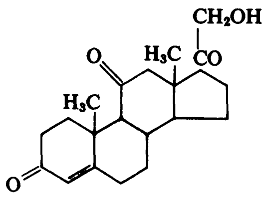
ADRENAL GLANDS. Part of the endocrine system, the adrenal glands are two small bodies located at the upper end of each kidney. The right adrenal gland is somewhat triangular in shape; the left gland is more semilunar. These glands range in size, but on the average weigh from 5 to 9 grams. Surrounding each gland is a thin capsule. There are two parts—the *cortex* (external tissue) and the *medulla* (chromophil tissue). These parts differ both in origin and function. In proportion to total body weight, the adrenals are much larger at birth and weight about 8 grams. In adults, proportionately, the adrenals are 1/20th the size they are in infants. Most of the shrinkage in size and weight occurs during the first year of life. The adrenals are well supplied with blood and nerves.

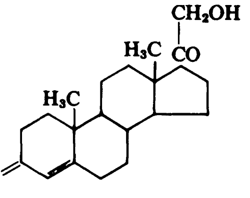
While the adrenal glands perform a number of functions, the two most important are: (1) Control of the body's adjustment to an upright posture; and (2) accommodation of the body to intermittent rather than constant intake of food. The adrenals also participate importantly in the regulation of electrolyte and water balance; the activity of lymphoid tissue and the number of eosinophils circulating in the blood (see **Blood**); the response to stress situations as may be encountered in infection, anesthesia, surgery, and volume loss; and the secretion of hormones. These hormones influence immune reactivity, blood cell formation, cerebral function, protein synthesis, and numerous other body processes. See also **Endocrine System; and Hormones.**

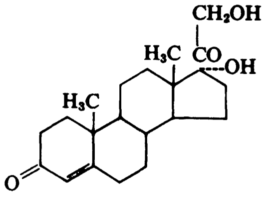
Several serious problems occur when there is imbalance or dysfunction of the adrenal glands, as, for example, in hyperfunction of the adrenal cortex, which precipitates Cushing's syndrome, of which there are several sub-forms; in hypofunction of the adrenal cortex; in adrenal neoplasms; and in pheochromocytoma.

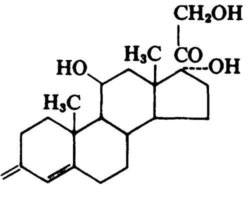
Adrenal Cortical Hormones. The hormones elaborated by the adrenal cortex are steroidal derivatives of cyclopentanoperhydrophenanthrene related to the sex hormones. The structural formulas of the important members of this group are shown in Fig. 1. With the exception of *aldosterone*, the compounds may be considered derivatives of *corticosterone*, the first of the series to be identified and named. The C_{21} steroids derived from the adrenal cortex and their metabolities are designated collectively as *corticosteroids*. They belong to two principal groups; (1) those processing an O or OH substituent at C_{11} (*corticosterone*) and an OH group at C_{17} (*cortisone* and *cortisol*) exert their chief action on organic metabolism and are designated as *glucocorticoids*; (2) those lacking the oxygenated group at C_{17} (*desoxycorticosterone* and *aldosterone*) act primarily on electrolyte and water metabolism and are designated as *mineralocorticoids*. In humans, the chief glucocorticoid is *cortisol*. The chief mineralocorticoid is *aldosterone*.

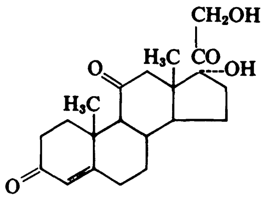

Corticosterone


11-Dehydrocorticosterone


Desoxycorticosterone


17-Hydroxy-desoxycorticosterone


17-Hydroxycorticosterone
(Cortisol)


17-Hydroxy-dehydrocorticosterone
(Cortisone)

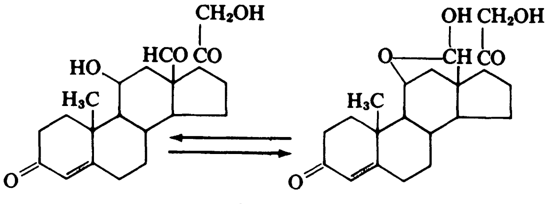

Aldosterone

Fig. 1. Adrenal cortical hormones.

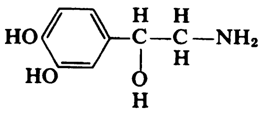
The *glucocorticoids* are concerned in organic metabolism and in the organism's response to stress. They accelerate the rate of catabolism (destructive metabolism) and inhibit the rate of anabolism (constructive metabolism) of protein. They also reduce the utilization of carbohydrate and increase the rate of gluconeogenesis (formation of glucose) from protein. They also exert a lipogenic as well as lipolytic action, potentiating the release of fatty acids from adipose tissue. In addition to these effects on the organic metabolism of the basic foodstuffs, the glucocorticoids affect the body's allergic, immune, inflammatory, antibody, anamnestic, and general responses of the organism to environmental disturbances. It is these reactions which are the basis for the wide use of the corticosteroids therapeutically. See also **Immune System and Immunology.**

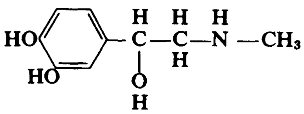
Aldosterone exerts its main action in controlling the water and electrolyte metabolism. Its presence is essential for the reabsorption of sodium by the renal tube, and it is the loss of salt and water which is responsible for the acute manifestations of adrenocortical insufficiency. The action of aldosterone is not limited to the kidney, but is manifested on the cells generally, this hormone affecting the distribution of sodium, potassium, water, and hydrogen ions between the cellular and extracellular fluids independently of its action on the kidney.

The differentiation in action of the glucocorticoids and the mineralocorticoids is not an absolute one. Aldosterone is about 500 times as effective as cortisol in its salt and water retaining activity, but is one-third as effective in its capacity to restore liver glycogen in the adrenalectomized animal. Cortisol in large doses, on the other hand, exerts a water and salt retaining action. Corticosterone is less active than cortisol as a glucocorticoid, but exerts a more pronounced mineralocorticoid action than does the latter. See also **Steroids.**

In addition to the aforementioned corticosteroidal hormones, the adrenal glands produce several oxysteroids and small amounts of testosterone and other androgens, estrogens, progesterone, and their metabolites.

Adrenal Medulla Hormones. Adrenaline (epinephrine) and its immediate biological precursor *noradrenaline (norepinephrine*, levarternol) are the principal hormones of the adult adrenal medulla. See Fig. 2. Some of the physiological effects produced by adrenaline are: contraction of the dilator muscle of the pupil of the eye (*mydriasis*); relaxation of the smooth muscle of the bronchi; constriction of most small blood vessels; dilation of some blood vessels, notably those in skeletal muscle; increase in heart rate and force of ventricular contraction; relaxation of the smooth muscle of the intestinal tract; and either contraction or relaxation, or both, of uterine smooth muscle. Electrical stimulation of appropriate sympathetic (*adrenergic*) nerves can produce all the aforementioned effects with exception of vasodilation in skeletal muscle.


Noradrenaline


Adrenaline

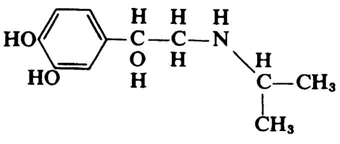

Isoproterenol (Synthetic)

Fig. 2. Adrenal medula hormones.

Noradrenaline, when administered, produces the same general effects as adrenaline, but is less potent. Isoproternol, a synthetic analogue of noradrenaline, is more potent than adrenaline in relaxing some smooth muscle, producing vasodilation and increasing the rate and force of cardiac contraction.

Malfunctions of the Adrenals. Progressive destruction of the adrenal cortex, as found in *Addison's disease*, gives rise to symptoms resulting from deficiencies of the cortical hormones previously mentioned. In the Waterhouse-Friderichsen syndrome, destruction of part or the whole of one or both glands by hemorrhage, as may occur in the course of meningitis due to meningococci, leads to sudden collapse and death unless very prompt treatment by replacement of the absent hormones is available. See also **Addison's Disease.** In *Cushing's syndrome* (adrenal cortical excess), a rare condition, there is obesity of the abdomen, face, and buttocks, but not of the limbs. The skin about the face and hands is redder than normal. Hair grows profusely, and women may grow excessive hair about the face. Bones become brittle and suffer a considerable loss of mineral components. Sexual functions may fall to a low level. The adrenal cortex may be overly stimulated by an excess secretion of ACTH (*adrenocorticotropic hormone*) in the pituitary gland. See also **Pituitary Gland.** Also, excessive production of ACTH may be caused by malignant tumors, among which lung, thymus, pancreas, and kidney are the most common. Similar symptoms also may be produced by the excessive pharmacologic (*iatrogenic*) use of steroids. Because of the multifaceted causes of Cushing's syndrome, differential diagnosis is important.

ADSORPTION INDICATORS. Dyestuffs or other chemicals which are used to detect the end point of a precipitation titration. These substances are dissolved in the solution to be titrated and lend color to it. The end point of the titration is signaled by disappearance of the color from the solution, or a change of color in the solution, attributable to adsorption of the indicator by the precipitate. See also **Indicator (Chemical).**

ADSORPTION (Process). The *physical* property of certain materials to adsorb certain other substances is used widely in the chemical and petrochemical industries for purifying (removing undesirable components) various gases, such as *syngases,* that are the feedstock for synthesizing numerous organic chemicals (methanol, for example) and also fuel gases (natural or synthetic). As a major chemical engineering *separations technique,* adsorption sometimes competes with absorption, q.v., membrane, and membrane separation operations.

Definition. Adsorption is a type of adhesion that takes place at the surface of a solid or a liquid in contact with another medium, resulting in an accumulation or increased concentration of molecules from that medium in the immediate vicinity of the surface. For example, if freshly heated charcoal is placed in an enclosure with ordinary air, a condensation of certain gases occurs upon it, resulting in a reduction of pressure; or if it is placed in a solution of unrefined sugar, some of the impurities are likewise adsorbed, and thus removed from the solution. Charcoal, when activated (i.e., freed from adsorbed matter by heating), is especially effective in adsorption, probably because of the great surface area presented by its porous structure. See Fig. 1. Its use in gas masks is dependent upon this fact. Penicillin is recovered in one state of the process by adsorption on activated carbon.

When colloidal hydroxides, notably aluminum hydroxide, are precipitated in a solution of acidic dyes, that is, those containing the groups —OH or —COOH, the dye adheres to the precipitate, yielding what is termed a lake. The "adsorption" of dirt on one's hands results from the unequal distribution of the dirt between the skin of the hands and the air or solid with which the skin comes in contact. Water is frequently ineffective in removing the dirt. The efficacy of soap in accomplishing its removal is due to the unequal distribution of dirt between skin and soap "solution," this time favoring the soap and leaving the hands clean.

At a given fixed temperature, there is a definite relation between the number of molecules adsorbed upon a surface and the pressure (if a gas) or the concentration (if a solution), which may be represented by an equation, or graphically by a curve called the *adsorption isotherm*.

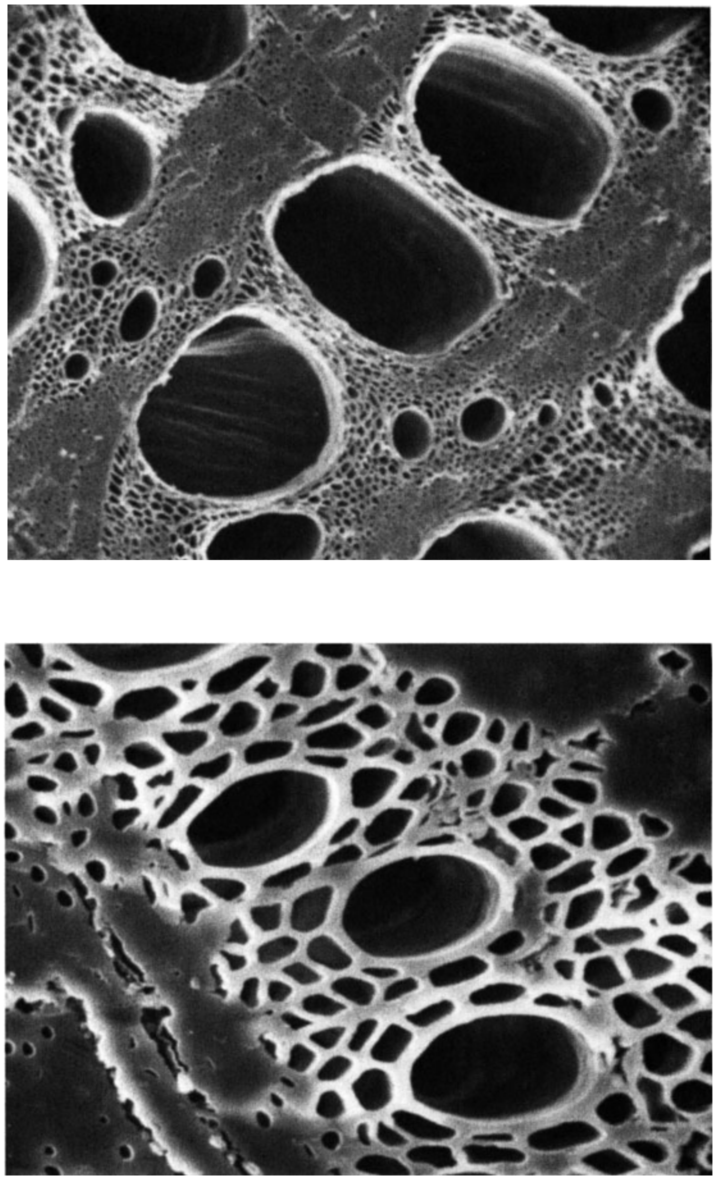


Fig. 1. Scanning electron micrographs of charcoal magnified 100× (top) and 50× (bottom). (*Polaroid Type 105 Land Film.)*

The Freundlich or classical adsorption isotherm is of the form:

$$\frac{x}{m} = kp^{1/n}$$

in which x is the mass of gas adsorbed, m is the mass of adsorbent, p is the gas pressure, and k and n are constants for the temperature and system. In certain systems, it is necessary to express this relationship as:

$$\frac{x}{m} = k(h\gamma)^{1/n}$$

where h is the relationship of the partial pressure of the vapor to its saturation value, and γ is the surface tension. Numerous isotherm equations have been proposed in the chemical literature in the last fifty years. The Langmuir adsorption isotherm is of the form:

$$\frac{x}{m} = \frac{k_1k_2p}{1 + k_1p}$$

The Brunauer, Emmett and Teller equation is more general than those of Freundlich or Langmuir; for among other limitations, those two equations apply only to the adsorption of gases. Even in those cases, the degree of adsorption depends upon five factors: (1) the composition of the adsorbing material, (2) the condition of the surface of the adsorbing

PHYSICAL PROPERTIES OF REPRESENTATIVE ADSORBENTS

Adsorbent Substance	Internal Porosity (%)	External Void Fraction (%)	Average Pore Diameter (10^{-10} meter)	Surface Area (square meters/gram)	Adsorptive Capacity (gram/gram of dry solid)
Activated alumina	25	49	34	250	0.14[a]
Activated bauxite	35	40	~50	—	0.04–0.2[b]
Fuller's earth	~54	40	—	130–250	—
Silica gel	~70	30–40	25–30	~320	1.0[c]
Shell-base carbon	~50	~37	20	800–1100	45
Wood-base carbon	55–75	~40	20–40	625–14,000	6–9[d]
Coal-base carbon	65–75	45–70	20–38	500–1200	~0.4[e]
Petroleum-base carbon	70–85	26–34	18–22	800–1100	0.6–0.7[f]
Anhydrous calcium sulfate	38	~45	—	—	0.1[g]
Synthetic zeolites—See Note below.—					

[a] Water at 60% relative humidity.
[b] Water; test condition not specified.
[c] Water at 100% relative humidity.
[d] Phenol value.
[e] Benzene at 20°C; 7.5 millimeters partial pressure.
[f] Test conditions not specified.
[g] Water; test conditions not specified.

Note: Synthetic zeolites are available from numerous commercial sources and usually are best known by their tradenames.

Pore Size: Ranges from 3Å (*Phillipsite*) to 6×7Å for *Mordenite*.
Composition: Si/Al: 1.0 for *US-Y* to 5–10 for *Ferrierite* and *Mordenite*.
Cation: Na, H, K, Ca, and Mg varies with different formulations.
Sorption Capacity (Wt %): H_2O: 1 for *Silicalite* to 28 for *Faujasiite X*
nC_6H_{14}: 1.3 for *Phillipsite* to 18.1 for *Faujasite Y*
C_6H_{12}: 1 for *Chabazite* to 19.5 for *Faujasite Y*
Note that, for all the foregoing, there are numerous intermediate values.

Suppliers include: Anaconda Minerals, Double Eagle Mining, W. R. Grace, Mobil Oil, PQ Corp, Toyo Soda, Union Carbide, et al. Adapted from data developed by D. E. W. Vaughan.

material, (3) the material to be adsorbed, (4) the temperature, and (5) the pressure (if a gas). A notable case in point is carbon. Of the finely divided varieties of carbon there are important sugar charcoal, bone black or animal black, blood charcoal, wood charcoal, coconut-shell charcoal, activated carbon. The temperature of preparation of adsorbent charcoal is an important factor, high temperatures being deleterious, and the removal (or non-removal) of gases by passing steam over the heated carbon, which operation increases the adsorptive power. Bone black is used for removing the coloring matter from raw sugar solutions. Fusel oil is removed from whiskey and poison gases from air by adsorption with the proper form of carbon. By cooling carbon in a vacuum to the temperature of liquid air, the concentration of residual gas is greatly decreased. Dewar (1906) found that 5 grams of charcoal (presumably coconut-shell charcoal) at the temperature of liquid air reduced the pressure of air in a 1-liter container from 1.7 to 0.00005 millimeters.

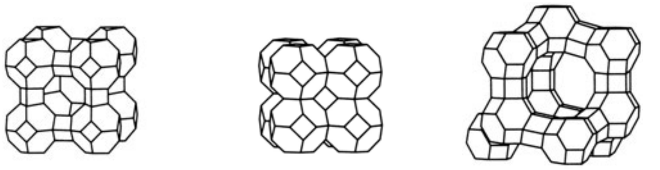
Major Adsorbents. Besides carbon, other important adsorbents are infusorial or diatomaceous earth fuller's earth, clay, activated silica (silica gel), activated alumina, and synthetic zeolites. See accompanying table. All surfaces that behave indifferently towards non-electrolytes have the ability to adsorb electrolytes.

As pointed out by Ruthven, the discovery and commercialization of synthetic zeolites as adsorbents provided the primary stimulus to the development of adsorption as a major separation process. Although, more recently, other materials, such as organic resins, carbon molecular sieves, and pillared clays, have appeared, the zeolites remain a dominant factor in adsorbent selection. See Fig. 2. Because of their microporous crystalline structure, zeolites are uniquely suited to the role of adsorbents. Although commercially available only in recent years, it is interesting to note that R. M. Barrow first recognized the adsorbent selectivity potential of zeolites in the 1940s. Earlier research was directed towards natural zeolites. (Synthetic zeolites also are used in catalytic processes. See **Catalysis.**)

Classes of Adsorbents. On the basis of the adsorbate (or the substance which is adsorbed), adsorption may be *polar*, when the material adsorbed consists of positive or negative ions, so that the adsorbed film has an overall electrical charge. The term polar adsorption is also applied to adsorption chiefly attributable to attraction between polar groups of adsorbate and adsorbent. *Specific* adsorption is the preferential adsorption of one substance over another, or the quantity of adsorbate held per unit area of adsorbent.

On the basis of the process involved, adsorption may be classified as *chemical adsorption* (or chemisorption) where forces of chemical or valence nature between adsorbate and adsorbent are involved; and *van der Waals adsorption*, involving chiefly van der Waals forces. The difference is usually indicated experimentally by the greater heat of adsorption and more specific nature of the chemical process.

In comparing the adsorption properties of silica gel and activated carbon it may be noted that the latter is nonpolar and since it has no



2 (a) 2 (b) 2 (c)

Fig. 2. Three types of zeolites used in adsorption operations. (a) A type, (b) *Sodalite,* and (c) *Faujasite* (Type X, Y). Zeolites are crystalline microporous solids that contain cavities and channels measured in terms of molecular dimensions. Sometimes called *molecular sieves,* these dimensions range from 3 to 10Å. At one time, zeolites were chemically classified as aluminosilicates. With the development of aluminophosphate chemistry, at least 13 additional elements have been added to their composition, including Li, Be, B, Mg, Co, Mn, Zn, P, As, and Ti, plus the earlier known substitution of Si, Al, Ga, Ge, and Fe. As pointed out by Vaughan, "Few fields of chemistry (inorganic) offer such chemical and structural diversity." There are 60 different known structures, but theorists suggest that tens of thousands may be possible. The foundation building block of zeolite structure is a tetrahedron of four oxygen atoms surrounding a central silicon atom (SiO_4). Natural zeolites also occur, q.v. (*Sketches adapted from Vaughan.*)

affinity for water will adsorb organic compounds in preference to water. Silica gel, on the other hand, is polar. It retains water and may thus reject an organic compound. It can also discriminate more selectively than activated carbon and consequently can be used for the fractionation of organic solvents.

Other types of adsorption are *oriented adsorption*, in which the adsorbed molecules or other entities are directionally arranged on the surface of the adsorbent; and *negative adsorption*, a phenomenon exhibited by certain solutions, in which the concentration of solute is less on the surface than in the body of the solution.

Occlusion is a type of adsorption, or perhaps more properly absorption, exhibited by metals or other solids toward gases, in which the gas is apparently incorporated in the crystal structure of the solid. Palladium thus occludes extraordinary quantities of hydrogen, with the simultaneous liberation of much heat.

Applications and Equipment Required. Industrial adsorption processes are relatively complex and consist of several steps, as in the case of removing (adsorbing) undesirable vapors and gases from a mixture of gases: (1) passing the gas mixture through an adsorption column under pressure, (2) regeneration of the column through the application of heat and a purging gas, (3) cooling the column, and (4) repressurization of the column to make ready for the next cycle.

Adsorption is extremely important in biological reactions. Many of the constituents of plant and animal cells are colloidal in nature, and materials of various sorts are adsorbed on the surfaces of these colloids. For example, proteins frequently depend upon hydration or the adsorption of water for their activity, and the cellulose walls of plant cells often adsorb much water. See also **Chromatography.**

During recent years, large installations for processing municipal wastes have been made. By taking advantage of anaerobic (do not require oxygen to live) bacteria, close to 100% of the waste mass can be converted to large volumes of usable gas (methane), leaving relatively small amounts of dry solids. The gas produced is known as landfill gas (LFG). Wells are drilled into the landfill, with the application of a slight vacuum to avoid drawing air into the base of the fill. LFG contains a variety of nonbiodegradeable chemicals, including chlorinated fluorinated hydrocarbons, plus the remainders of hundreds of packaged chemicals (solvents, cleaners, et al.). The raw LFG also contains large quantities of carbon dioxide, which must be removed. Nearly all of the undesired components can be removed via selective adsorption. Usually, the target is to produce a medium-energy gas, for use as boiler fuel or in gas turbine electrical generators, or a high-energy, pipeline-quality methane stream.

As of 1990, it is estimated that existing or proposed landfill plants will be producing well over 100 million SCF (standard cubic foot), equivalent to nearly 3.5 billion cubic meters of raw LFG per year.

Adsorption plays an important role in the process of dyeing, and in contact catalytic processes such as the conversion of sulfur dioxide to trioxide, and of nitrogen plus hydrogen to ammonia. In the case of insoluble organic acids (containing —COOH group) and substances containing hydroxyl (—OH) groups on the surface of water, the film is oriented so that the —COOH or —OH groups are attracted into the surface of the water, while their hydrocarbon ends project away from the surface of the water showing no tendency to dissolve (Langmuir).

The heat of adsorption, or wetting in this case, of starch by water is 29 calories per gram of dry starch. The heat of adsorption of various vapors and adsorbents has been measured. Since increase of temperature reduces adsorption, the adsorption process is accompanied by the evolution of heat. It appears that the heat liberated for a given volume of liquid filling the capillary spaces of a given adsorbent is practically constant. The heat of adsorption of hydrogen is, on nickel, palladium, platinum, copper, 11,700, 18,000, 13,800, 9500 calories respectively per gram mol (2 grams) of hydrogen; and of carbon monoxide on platinum 35,000 calories per gram mol (28 grams) of carbon monoxide; and of ethylene on copper 9500 calories per gram mol (28 grams) of ethylene.

Additional Reading

Barrer, R. M.: "Zeolites snd Clay Minerals as Molecular Sieves," Academic Press, London, 1978. (A classic reference.)

Chen, N. Y., and T. F. Degnan: "Industrial Catalytic Applications of Zeolites," *Chem. Eng. Prog.*, 32 (February 1988).

Gubbins, K. E.: "Molecular Adsorption in Micropores," *Chem. Eng. Prog.*, 42–44 (August 1990).

Kumar, R., and J. K. Van Sloun: "Purification by Adsorptive Separation," *Chem. Eng. Prog.*, 34–40 (January 1989).

LeVan, D., Editor: "Adsorption and Ion Exchange Fundamentals and Applications," Amer. Socy. of Chem. Engrs., New York, 1988.

Liapis, A. I., Editor: "Fundamentals of Adsorption," Amer. Socy. of Chem. Engrs., New York, 1987.

Ma, Y. H., and J. P. Ausikaitis, Editors: "Recent Progress in Adsorption and Ion Exchange," Amer. Socy. of Chem. Engrs., New York, 1987.

Ruthven, D. M.: "Zeolites as Selective Adsorbents," *Chem. Eng. Prog.*, 42–50 (February 1988).

Staff: "Gas Process Handbook '92," in *Hydrocarbon Processing*, 85 (April 1992).

Vaughan, D. E. W.: "The Synthesis and Manufacture of Zeolites," *Chem. Eng. Prog.*, 25 (February 1988).

White, D. H., Jr., and P. G. Barkley: "The Design of Pressure Swing Adsorption Systems," *Chem. Eng. Prog.*, 34 (January 1989).

Yang, R. T.: "Gas Separation by Adsorption Processes," Butterworths, New York, 1987.

ADVENTITIOUS BUDS. Buds which appear elsewhere than in the leaf axils or above them. They may appear anywhere in the internode, or on roots or even on leaves, and develop either naturally or as a result of injury. The dense bunches of buds, which frequently appear on burls or in witches'-broom, may be adventitious. The buds which appear at the tops of thistle roots, particularly when the natural top of the plant is cut off, are adventitious. So also are the buds which develop on the leaves of the begonia and bryophyllum. The practice of pollarding, or cutting off the branches of a tree in such a way as to leave only the main trunk or perhaps the stumps of a few large branches, results in the development of dense groups of adventitious buds. These grow into adventitious branches which in certain willows may be long and supple and so useful in manufacturing wicker furniture. The habit of forming adventitious buds on roots is of material value, since in consequence it is possible to propagate many plants by means of root cuttings.

Adventitious roots also exist. They may appear from the stem where they arise in the pericycle, or from other tissues of the plant. The roots which appear on slips or stem cuttings are adventitious. See also **Bud;** and **Budding.**

AEOLIAN TONES. The tones produced by a gas stream striking a stretched wire in a direction normal to the length of the wire.

AERATION. A process of contacting a liquid with air, often for the purpose of releasing other dissolved gases, or for increasing the quantity of oxygen dissolved in the liquid. Aeration is commonly used to remove obnoxious odors or disagreeable tastes from raw water. The principle of aeration is also used in the treatment of sewage by a method known as the activated sludge process. The sewage is allowed to flow into an aeration tank where it is mixed with a predetermined volume of sludge. Compressed air is introduced which agitates the mixture and furnishes oxygen which is necessary for certain biological changes which take place. Sewage may also be aerated by mechanically actuated paddles which rotate the liquid and constantly bring a fresh surface in contact with the atmosphere.

Aeration is of importance in the fermentation industries. In the manufacture of baker's yeast, penicillin, and other antibiotics, an adequate air supply is required for optimum yields in certain submerged fermentation processes.

Aeration can be accomplished by allowing the liquid to fall in a thin film or sprayed in the form of droplets in air at atmospheric pressure; or the air, under pressure, may be bubbled into the liquid as by means of a sparger, or other device that creates thousands of small bubbles, thus providing maximum contact area between the air and the liquid.

AERENCHYMA. Spongy tissue occurring chiefly in the stems of many aquatic or marsh plants. Such porous tissue gives great buoyancy to the stem and so helps keep the leaves up in the air and permits gas diffusion within the plant. The term aerenchyma is frequently applied to any loose porous tissue found in plants, as for example that occurring in lenticels.

AEROBE. An organism that utilizes atmospheric oxygen in its metabolic processes; i.e., the so-called aerobic bacteria, or aerobes, which use oxygen. See also **Anaerobe.**

AERODYNAMICS AND AEROSTATICS. *Aerodynamics* is the study of the *interaction between* (1) a *fluid*, most frequently air, that is stationary or in motion (wind), and (2) a *solid object* that (a) is essentially *stationary* (fixed), such as a building, chimney, bridge, or that (b) is in *motion*, such as an airborne or landborne vehicle, a missile, a sporting object (ball, boomerang), et al. This interaction, when translated to practical engineering terms, is of great importance in terms of designing structures and vehicles for maximum safety, performance, and efficiency. For example, in terms of fixed structures, a bridge must be designed to minimize the effects of high wind—to avoid undue swaying and possibly toppling under extreme conditions. (Bridges have been destroyed—e.g., the Tacoma Narrows Bridge, Washington, 1940.) For safety and other reasons, highrise buildings must be designed to minimize the effects of wind. Aerodynamics is probably most commonly associated with the design of all types of aircraft, but much attention is also given to landcraft, such as automobiles, trucks, and even bicycles. The interest in drag, discussed later, as it affects land vehicles has been intensified since the energy crisis of a few years ago—because a significant consumption of fuel is consumed in overcoming drag. The aerodynamical characteristics of vehicles also affect performance (the smoothness and grace of operation) as well as efficiency; also safety, particularly at high speeds, can be affected by aerodynamically designing an auto or truck that "holds the road" so to speak. Aerodynamics is or should be a major consideration in the design of trailing vehicles so that they can be operated safely under windy conditions. More recently, the disciplines of aerodynamics and acoustics have been combined to tackle "wind noise" in passenger cars. See **Acoustics.** The first streamlined train appeared in the 1930s (the so-called zephyrs) and aerodynamics was of major concern in the design of today's fastest trains, such as those operating in France and Japan. *Streamlining* is a general lay term for designing vehicles that incorporate sound aerodynamic engineering and that, aesthetically, appear to be so designed. There have been cases in the past where a so-called streamlined "look" was effected much more for buyer appeal than for the achievement of aerodynamic performance.

Aerodynamics is a complex, essentially mathematical and geometric science, but also one that has learned much from nature, rediscovering as it were the streamlining of birds, insects, and fast-moving land mammals. Aerodynamics depends upon the results of massive data collection, ranging from model testing (sometimes in huge wind tunnels) to test piloting aircraft and road testing of land vehicles. As will be touched upon later, computer simulation has aided and, in some cases, replaced the need for costly wind tunnel procedures.

Aerostatics. The basic principle of aerostatics can be attributed to Archimedes—a body immersed in a fluid (liquid or gas) is buoyed up by a force equal to the weight of the fluid displaced. Thus, for buoyancy, the weight of a structure (balloon, dirigible, etc.) plus the weight of the contained gas must equal the weight of the air displaced.

Fundamentals of Aerodynamics

Concept of Inversion. Whether an object is standing still and the air around it is in motion, or whether an object is moving in still air, provided the velocity of the object in the one case is the same as the velocity of the air in the other case, the aerodynamic forces prevailing are the same. This concept of inversion is exploited when determining the aerodynamic forces created by airstreams upon stationary airfoils and other objects as they may be tested in the moving airstreams of a wind tunnel. More simply stated, the wind detected by a person standing in a boat that is moving at 10 miles per hour, aerodynamically speaking, is the same as the wind felt by a person standing in a fixed position with a wind of 10 mph blowing.

Properties of Air. Air has a viscosity of 0.00019 poise under standard conditions, compared with 0.01 poise for water and 1 poise for light oil. Thus, there are fields of aerodynamic study in which the air can be considered frictionless, especially in subsonic aerodynamics—with the exception of the conditions existing in a very thin layer immediately adjacent to the solid object where there is relative motion between the object and the air. However, at speeds greater than that of sound, friction effects become important. In aerodynamic expressions, the absolute viscosity, μ, usually is associated with mass density, ρ, as a ratio. This is called the *kinematic viscosity*, ν, which has a magnitude of 0.000157 ft^2 (0.0000146 m^2) per second in standard sea level atmosphere.

At subsonic air speeds, it is customary to consider the air incompressible, and to rely on Bernoulli's theorem of the interchangeability of static head and velocity head. This permits the employment of the principle of continuity for flow, represented by the equation

$$AV = \text{constant}$$

which means, symbolically, that the product of the cross-sectional area of flow and the velocity is a constant at all points in the path taken by a finite quantity of the fluid. If the fluid were compressible, this continuity would be

$$\rho AV = \text{constant}$$

Streamtube. A streamtube in air may be visualized as an imaginary conduit through which the condition of continuity of flow of the incompressible fluid takes place. Although the flow is steady, the tube is not necessarily of uniform size, but where its sectional area decreases, we would anticipate an increase of velocity; furthermore, a consideration of Bernoulli's theorem would indicate that these diminished sections of the streamtube are regions of lower static pressure than elsewhere. Although a streamtube is ordinarily imagined to have a circular cross section, it will be convenient here to think of a streamtube having a rectangular cross section. If we have enough of these streamtubes, all of the same width, but whose heights may vary as velocity varies, we might imagine them stacked, one on another, as shown in Fig. 1, and be thus encompassing the whole flow of air through a given region. It will be convenient to think of the edges of these streamtubes as equivalent to the streamline pattern that is so useful in the study of the effect of airflows past solid objects. It follows that where the streamlines approach each other, there is a region of lower pressure and higher velocity; conversely when they diverge. Thus, converging streamtubes (and streamlines) are like nozzles, whereas diverging streamtubes are like diffusers. The reader, upon viewing a streamline pattern, may then judge fluid pressure by the density of streamlines. This is readily done, qualitatively, if not quantitatively.

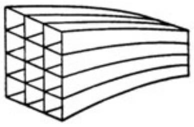


Fig. 1. Stacked streamtubes.

Streamline Patterns. It is now in order to state the nature of elementary streamline patterns and explain how a complex streamline pattern may be considered the result of the addition of two potential forces, each capable of producing a different character of streamline in the same region. A simple potential cylinder rotating in still air will drag with it a circulatory pattern of flow, the streamlines of which could be represented by a series of concentric circles. The circulation stream pattern is of great importance, because an airfoil is a shape that automatically induces a circulating component of airflow, which, when superimposed upon a rectilinear stream, provides the typical streamline pattern around an airfoil. The lifting streamline pattern of the typical airfoil could be induced in the absence of the airfoil by imposing some specific circulatory pattern of stream flow upon a rectilinear field, and this fact has been put to useful service in theoretical mathematical consideration of the aerodynamics of airfoils.

In Fig. 2a, a rectilinear flow is passing a static cylinder. At some distance from the cylinder, the flow is unaffected by the presence of the cylinder, and the streamtubes remain of constant size throughout, but it is obvious that between this undisturbed region and the cylinder, the intermediate tubes must be smaller in order that all can be accommodated in the space immediately above and below the cylinder. The action of the individual streamtube approaching a point abreast of the

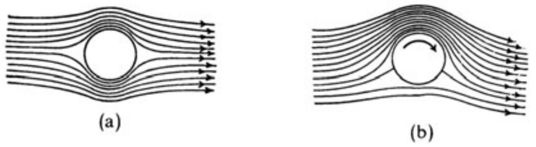


Fig. 2. Comparison of air passing over (a) cylinder at rest and (b) rotating cylinder.

center line of the cylinder is, therefore, that of a nozzle which has increased the air speed and decreased its pressure. Hence, above and below the cylinder, static pressure is less than in the free, unaffected stream. If the cylinder were in rotation, tending to set up its concentric field of streamlines, the resulting streamline pattern would be like that shown in Fig. 2b, where, on one side of the cylinder, there is a reduction of static pressure due to crowding together of the streamlines, and on the other side, there may be no change of pressure, or even an increase. Briefly, then, the presence of a circulatory component of flow in a rectilinear field is to produce a net transverse force (in this case, upwards) on the object, since it will be subjected to unbalanced static pressure of the fluid against its surfaces. This is the force which causes a spinning tennis ball to assume a sharply curved trajectory, and which causes the lift to exist upon an airplane wing.

Effects of Viscosity. The influence of viscosity of air in aerodynamics is confined chiefly to the action of air in the boundary layer. Within this boundary layer (which, although extremely thin, constitutes the atmosphere immediately adjacent to the solid) the frictional qualities of air are of importance in determining the airflow. The thin boundary layer might be imagined subdivided into a great many laminae of air parallel to the surface and extending out to the limits of the boundary layer. A nonturbulent boundary layer is considered, although, as will be developed later, the boundary layer may become turbulent under certain conditions except for the lamina immediately adjacent to the stationary surface. As Fig. 3 displays, the laminar velocity, u, increases above the surface until, at the edge of the boundary layer, it equals the free stream velocity. The smaller the viscosity of the fluid, the thinner the boundary layer will be. Consequently, it is extremely thin for air, but what goes on in this boundary layer is of utmost importance in aerodynamics. The first layer of air sticks or adheres to the surface, and laminae above it successively slide on each other, exerting drags that are proportional to the viscosity. The rate of change of the velocity between adjacent laminae is a measure of the unit shearing force between them, and is the unit skin friction when the lamina considered is at the one in motion nearest the stationary surface. This occurs at practically zero boundary layer thickness. A curve joining the tips of velocity vectors plotted for the different laminae in the boundary layer is called a velocity profile. If u is the variable velocity, increasing from 0 to V across the boundary layer, then du/dy is the rate of change of velocity (also the tangent to the velocity profile) and, through the application of one of the Newtonian principles, the unit shear stress,

$$\tau = \mu\left(\frac{du}{dy}\right)_{y=0}$$

where y is measured in the vertical direction.

Surface Effects. The proximity of streamlines to a stationary surface is now seen to produce a reaction resulting from the static pressure p existing in the flow nearest the surface, and a skin friction τ, which is, of course, tangential to the surface. The combination of these two is a force oblique to the surface. The foregoing statement premises an adherence of the streamlines to the surface, for, if they have separated from the surface, the region between will be filled with vortices of random motion, and the surface pressure cannot be definitely predicted from the character of the streamline flow. It is apparent that aerodynamic reaction is a function of μ, ρ, V, and the extent of the surface area on which the reaction takes place. Usually some one dimension l is taken to describe the size of the surface. In ft-lb sec units, the aerodynamic force F is in pounds, μ is in poises, ρ is in slugs per square foot, and l is in feet. To discover the relationship existing between the reaction and its controlling factors, it is customary to set up the dimensional relationship:

$$|F| = |\rho^a \mu^b V^c l^d|$$

and employ the methods of dimensional analysis to discover the value of a, b, c, and d. It is found that

$$|F| = \left|\left(\frac{\nu}{Vl}\right)^b \rho V^2 l^2\right| \quad \text{dimensionally}$$

and

$$F = K\left(\frac{\nu}{Vl}\right)^b \rho V^2 l^2 \quad \text{numerically}$$

in which

ν = kinematic viscosity
K = a dimensionless constant
Vl/ν = the Reynolds number (also dimensionless)

The product, $2K(\nu/Vl)^b$, is frequently given as a dimensionless coefficient C. The exponent b is of such magnitude that the Reynolds number has a minor (but not negligible) effect on C. The coefficient is affected primarily by the attitude of the body in the airstream. If it is completely symmetrical, as is a sphere, this variation is nonexistent, and C is a true constant except for the above-mentioned minor effect of the Reynolds number variation. However, the nonsymmetrical shape of an airfoil causes large and typical variations of C, depending on the attitude with which the airfoil is presented to the airstream. The force F would then be $C(\rho V^2/2)S$, or CqS, in which q is dynamic pressure, and S is some significant surface area, in square feet, of the object.

Laminar and Turbulent Flow. Flow within a streamtube may be laminar or slightly turbulent. It was stated that, in laminar flow, drag per unit area is $\mu(\partial u/\partial y)_{y=0}$, but turbulent flow produces higher skin friction because the velocity profile is fuller close to the surface due to turbulent energy interchange between laminae. Although at the beginning of contact between airstream and surface, true laminar flow existed, it has been found that a certain critical value of Reynolds number, determined chiefly by the linear dimension of contact passed over, the boundary layer suddenly becomes turbulent, and thickens. In spite of the higher skin friction, the drag on bluff bodies may actually become less with turbulence, because of an action within the turbulent boundary layer that resists separation of the streamlines from the surface. Viscous drag gradually slows down the fluid in a laminar boundary layer, and if contact with the surface is maintained over sufficient length, the innermost parts of the boundary layer will be brought to rest. This becomes the separation point. If a negative pressure gradient exists beyond the separation point, as it often does, there will be a reverse flow toward the separation point, all of which will produce a breaking away of the streamlines from the surface, leaving the surface between the two separation points, on either side of body, in contact with air of random vortices and low pressure. See Figures 4 and 5. It can be seen that because of the vortices, the pressure at n has failed to build up to the pressure at m (as it would have had the fluid been frictionless and without premature separation point), resulting in the creation of a downstream dragging force. This is characteristic of any body having a turbulent wake. Early turbulence in the boundary layer is desirable, so that

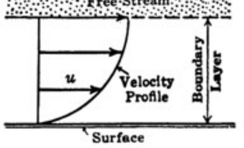


Fig. 3. Velocity profile.

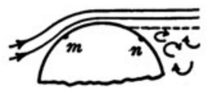


Fig. 4. Low-pressure wake.

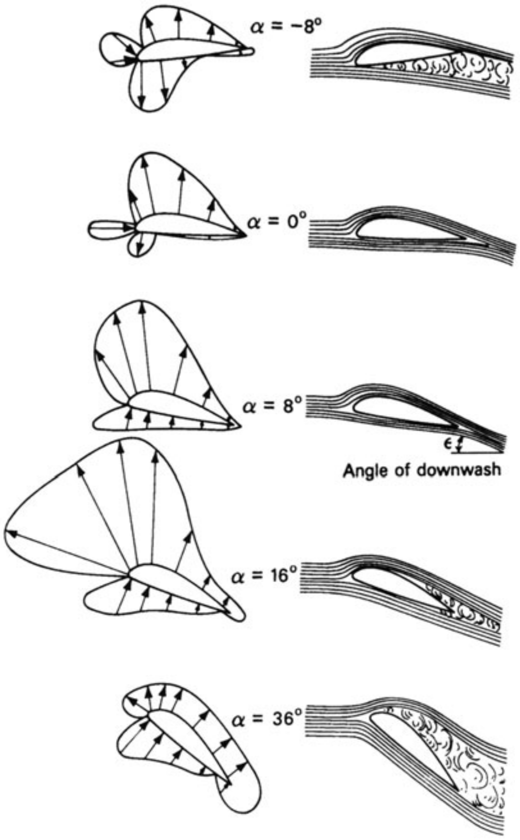


Fig. 5. Distribution of pressure on a cambered wing at different angles of attack but at the same airspeed as obtained in a wind tunnel. The arrows are force vectors, their length representing the relative intensities of the pressures at the points at which they are located. The end points (arrow-heads) are connected to form an envelope. To obtain the relative intensity at any point, erect an arrow normal to the airfoil surface, and its length is then determined by its intersection with the envelope drawn through the end points of the other force vectors. Sketches to the right show, approximately, the manner of air flow at the different angles of attack.

the greater average momentum of boundary layer air can carry it against skin friction farther around the bluff surface, and so delay the separation point and narrow the turbulent wake. Thus, the coefficient of drag for a sphere [C of $Cq(\pi d^2/4)$] is about 0.5 for laminar boundary layer and 0.1 for turbulent, with the transition occurring at a Reynolds number of approximately 300,000.

Airfoils

An airfoil is any body whose shape causes it to receive a useful reaction from an airstream moving relative to it. This definition is broad and could include many shapes not ordinarily considered airfoils. The term is usually associated with a body of the shape shown in Fig. 6, a profile or chordwise section of the wing. The dimension perpendicular to this section is called the span when referred to the wing. Many different airfoil shapes have been used, proposed for use, or tested. Some of them have flat lower surfaces, and others have convex or concave surfaces.

The characteristic lifting airfoil profile has a maximum thickness of 6 to 18% of the chord at 20 to 40% of the chord aft of the leading edge. Early experiments were made with flat-plate airfoils, and later with thin curved-plate airfoils, but neither possessed as large a ratio of lift to drag at a particular angle of attack as the double-surface cambered airfoil shown, nor as low resistance or drag. Also, an airfoil of finite thickness provides space for the foundation structure of a lightweight wing, thus removing the structural elements from the drag of the airstream, and thereby reducing the drag.

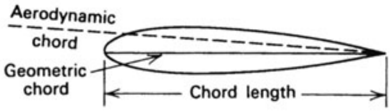


Fig. 6. Airfoil profile.

The value of the airfoil shape to aviation resides in the magnitude and direction of the resultant air reaction on it when employed at attitudes below the stall (about 15° angle of attack). The component of the air reaction normal (called the lift) to the free airstream is several times the magnitude of the parallel component (called the *drag*). The normal component, or lift, may be expressed in equation form as

$$L = C_L q S$$

The variation of C_L with geometric angle of attack is seen in Fig. 7. The slope of the straight portion is a per radian. Theory indicates a value of 2π for a, for infinite aspect ratio, and tests agree well with the theory. The lift coefficient increases uniformly with attack until the stall or burble point is reached, where it breaks and decreases rapidly with further increase of angle of attack, because the high attack angle creates a bluff body and the stagnation point advances forward on the airfoil. Normal controlled flight is not likely beyond the stall or burble point. The moment coefficient, C_{M_a}, about the aerodynamic center of the airfoil is shown in Fig. 7. It will be noted that it has practically a constant value. When listed, the value of C_{M_a} at $C_L = 0$ is usually given.

The full aerodynamic reaction upon an airfoil is a force of which the lift is the component normal to the free stream velocity. The other component is measured in the direction of the air velocity and is called drag. Drag coefficient has a characteristic parabolic variation, as Fig. 7 shows. The drag equation is analogous to that for the lift:

$$D = C_D q S$$

If the resultant of all air reactions on an airfoil is consolidated into a single imaginary force, it must act at the center of pressure of the airfoil. The figure shows that the location of this center of pressure of an airfoil varies with its attitude. The qualities of an airfoil that are sought for when employed in aviation are:

(1) High maximum C_L, in order to give low landing speed for a given size wing and weight.

(2) Low minimum C_D, so that the high speed, which occurs at small angles of attack, may be the greatest possible.

(3) High ratio of C_L to C_D, so that an efficient, economical airplane will result.

(4) A shape well suited to the construction of a strong, but lightweight, wing, at minimum cost.

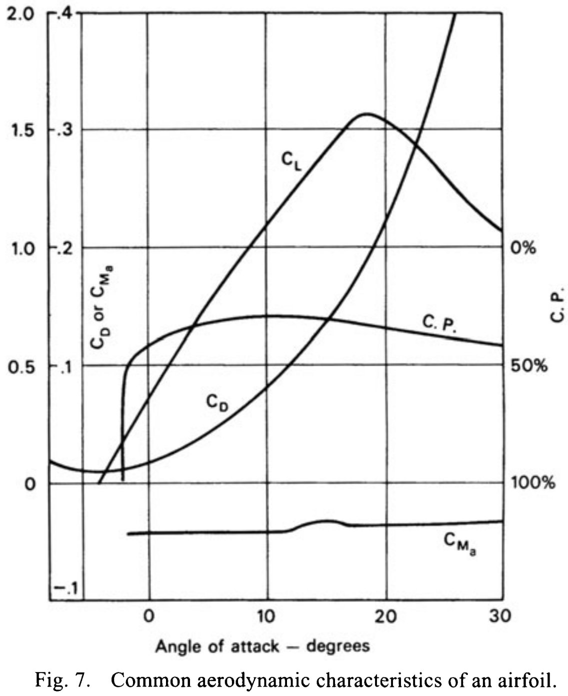


Fig. 7. Common aerodynamic characteristics of an airfoil.

Downwash. From an airfoil of finite span in flight, there is a multitude of vortices trailing, forming a vortex sheet. As a result of circulatory flow about each vortex filament a net downwash flow of air is produced, having an important effect upon the air reactions. If the airfoil is untwisted, and elliptical in plan form (or tapered in approximation of an ellipse), the vortices bound to the airfoil will produce a circulation of approximately elliptical strength distribution over the span. This produces a uniform downwash of air in the region of the wing. The combined effect of the forward motion of the airfoil and the downwash will be better understood upon reference to Fig. 8.

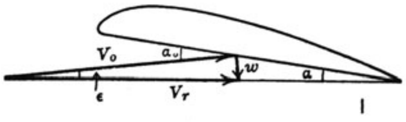


Fig. 8. Downwash.

The forward speed V_0 and the vortex-induced downwash w produce a resultant air velocity of V_r (substantially equal to V_0 in magnitude) which meets the airfoil at angle of attack α. The angle ϵ is called the downwash angle. Without downwash, the lift produced would be that corresponding to α_0 at velocity V_0, but downwash reduces the angle of attack to α thus decreasing the lift. To lift the same this airfoil would require α with downwash to be larger than α_0 without downwash. The effect of downwash is now seen to be deleterious since, for the same V and lift, it increases the needed α and produces extra drag. Furthermore this drag (induced drag) is not associated with aerodynamic cleanliness but with aspect ratio. Disposing the area of the wing as a slender wing of large wing span (i.e., high aspect ratio) is the way to reduce this form of drag.

The effect of downwash of the wing has to be accounted for in the disposition of the empennage (stabilizing tail assembly) when disposed in the rear of the wing. In front of the wing, there is an upwash.

High-Lift Devices. The lifting ability of an airfoil is improved by increasing its camber (over limited range), and also by delaying the separation of the air flow from the lifting surface as the angle of attack is increased. Most airfoils employed as airplane wings are of low camber because their drag coefficients at low attack angles (high speed) are more favorable than those of the thick, highly cambered airfoils.

It is known that the maximum lift coefficient of a highly cambered airfoil is greater than that of a flat, thin one. Safety in landing an airplane demands low landing speed, which is obtainable by using wings with high maximum lift coefficients. On the other hand, high maximum speed is only to be obtained with the thin, low-drag type airfoil. Many manufacturers of aircraft have adopted the flap as a means for securing low landing speeds on aircraft fitted with inherently high-speed wings. The flap, as illustrated in Fig. 9, is akin to a flexible trailing edge, and, by depressing the flap, the effective camber of the airfoil may be increased. The drag is also increased, giving a steeper gliding angle with flap down than without the use of the flap. Apart from any disadvantage, this factor may actually be beneficial, since with flap position variable and under control of the pilot, the angle of normal gliding is alterable, permitting steep approaches to small obstructed fields. Flaps that have received a great deal of attention include the simple flap, the split flap, zap flap, and the fowler flap. The simple flap, of which the normal aileron, elevator and rudder are types, is the easiest to use, requiring only rotation downward for increasing the lift coefficient. The split flap is extensively used. The operation is simply that of a rotation of a panel on the underside of the trailing edge about its forward edge. In the zap flap, the rotation is accompanied by a backward movement of the pivoting point. In the fowler flap, the motion is not one of simple rotation alone, but is combined with a rearward movement of the flap area as well. The effect is one of increase of airfoil area as well as change of camber.

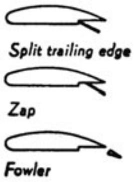


Fig. 9. Types of flaps.

Increasing lift by delaying separation in the boundary layer is an independent action. Both flaps and partial boundary layer control with the aid of a leading edge slot (see Fig. 10) could be added to the same airfoil—and have been on some airplane wings. The effect of flaps and slots on lift and drag is illustrated in Fig. 11. Delay of airflow separation by sucking off the incipient vortices around the stagnation point produced amazing lift coefficients but has not been exploited on account of the compressor energy needed, although, with the introduction of the jet engine, the interest in boundary layer control was revived. Also, very small high-speed jets to impart energy to the boundary layer, "blowing off" the air (as opposed to sucking off), and rotating cylindrical nose sections, have been tested with varying results.

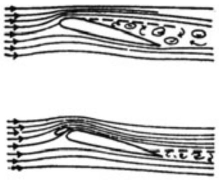


Fig. 10. Air flow at high angle of attack with and without slot.

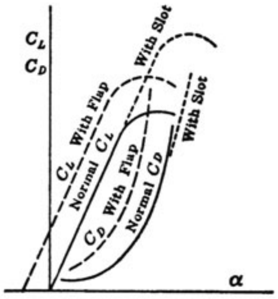


Fig. 11. Effect of flaps and slots on lift and drag.

Boundary layer control by leading edge slots is definitely valuable for lifting airfoils, and lacks the structural and mechanical complications attending the other boundary layer control proposals. Hence slots and flaps have been the principal high-lift additions commercially employed.

Drag

An object subjected to an airstream is acted upon by a resultant air pressure. Drag is the component of air reaction which is parallel to the airstream. Its origin is profile impact of molecules of air against the face of the object, the skin friction of the molecules of air as they slide along the object, the vortices and eddying air currents set up in an otherwise undisturbed airstream by the presence of the object, and the induction effects of downwash, if any. Drag is that quantity which imposes limitations upon the top speed of vehicles. As it is proportional to the square of the velocity, its magnitude mounts rapidly as velocities are increased.

There are two kinds of drag—drag on surfaces which obtain a useful reaction from the airstream as well as a drag, and drag upon surfaces where the only reaction is the drag. A wing has both drag and lift. Drag on the wing is the price paid for lift, and is so accepted. A strut, or wire, or wheel, creates no lift, and the drag is wholly undesirable. A drag of this type is called parasite drag. Careful streamlining and reduction of parts exposed to the airstream are ways of reducing parasite drag. The drag of a wing may be divided into profile drag and induced drag. Induced drag depends upon the lift coefficient and the aspect ratio of a wing. The induced drag coefficient may be expressed as

$$C_{D_i} = C_L^2/\pi(\text{AR})$$

where C_{D_i} is the induced drag coefficient, C_L the lift coefficient, and AR the aspect ratio of the wing.

During a test, profile and induced drag are not separable. Parasite drag depends upon the surface roughness and shape of the object.

Profile drag of airfoils is the result of skin friction and turbulent wake. It is greatly influenced by the Reynolds number (scale effect) and by initial turbulence in the airstream. Although some authors have proceeded on the basis that this is the sole influence on profile drag coefficient, others have found that the profile drag coefficient has a small variation with coefficient of lift. However, the generally accepted theories of origin of aerodynamic forces imply that scale effects modify profile and parasitic drags only and are absent in lift and induced drag. Experience confirms the validity of theory except near the burble, or stall, attitude of the airfoil, where viscosity has an effect on lift.

Experiments in the wind tunnel show that factors affecting reaction of air on airfoils are:

(1) The relative velocity of the air and airfoil.
(2) Ratio of span to chord of the surface area.
(3) Density of the air.
(4) The angle of inclination of the airfoil to the airstream.

All this may be stated somewhat as follows:

$$F \sim R^n \alpha \rho S V^2$$

F is wind reaction, α the angle of attack of the wing, ρ the mass density of the air, S the surface area, V the air velocity, and R the Reynolds number.

If a proper constant K be inserted, the similarity can be made into an equality. K may also be made to include the effect of the Reynolds number, angle of attack, and air density. The equation is then simplified to:

$$F = KSV^2$$

This reaction is neither perpendicular nor parallel to the wind stream. It is convenient to divide it into its components of lift and drag. Letting K_L and K_D be the corresponding coefficients.

$$\text{Lift } L = K_L S V^2$$
$$\text{Drag } D = K_D S V^2$$

There will be a different value of K for each angle of attack and each R, although the effect of the latter is usually minor.

The defect of this equation is that K is not dimensionless, as a true coefficient should be; also, the formula is not flexible with respect to air density. If it were rewritten:

$$R = C \rho S V^2$$

C would be dimensionless. The equation is usually written thus for drag:

$$D = \tfrac{1}{2} C_D \rho S V^2$$

because $\rho V^2/2$ is the pressure necessary to give air of mass density ρ at a velocity of V. This is called the *dynamic pressure*. The drag is also $C_D q S$, where q stands for dynamic pressure. The drag coefficient C_D varies with angle of attack, but it is dimensionless. Consistent units have to be used. ρ is the ratio of the weight of the air, or fluid, in pounds per cubic foot, to the acceleration due to gravity or 32.2 ft/sec^2 (9.8 m/sec^2); S is the wing area, projected upon the plane of chords, in square feet, and V is the velocity in feet per second. C_D is the nondimensional drag coefficient. The subscript D is used when the coefficient refers to the *average* coefficient for the *entire* wing, and the lower case letter d is used when the drag coefficient refers to the airfoil of unit width. The wind-tunnel test is used to establish the law of variation, and the results are plotted as a drag curve, with the coefficient of drag as the ordinate, and either coefficient of lift or angle of attack as the abscissa.

A few typical coefficients of drag are:

(1) Flat plate normal to airstream. Average value 1.28 often used, although subject to variation from AR (aspect ratio) and R.N. (Reynolds number).

Flat plate parallel to airstream. C_D = skin friction coefficient as there is no turbulent wake. Varies, 0.002–0.008 depending on initial turbulence and R.N.

(2) Sphere (S = projected area), about 4×10^5 critical R.N. Laminar boundary, 0.45; turbulent, 0.1.

(3) Cylinder, axis transverse to airstream (S = projected area), about 4×10^5 critical R.N. Laminar boundary, 0.7; turbulent, 0.3.

(4) Thin wing, infinite AR, R.N. 3.5×10^6. $C_{D_p} = 0.01$.

(5) Airplane. $C_D = C_{D_o} + C_L^2/C_{D_p}$ varies greatly, as it depends upon a large number of design and operational factors.

The drag of an object is subject to other factors also, such as compressibility of the fluid medium in which the object moves, and surface roughness.

Under supersonic conditions, the drag coefficient may be approximated by the following formula,

$$C_D = \frac{4\alpha^2}{\sqrt{M^2 - 1}}$$

where α is the angle of attack in radians, and M is the Mach number for the fluid speed considered.

Drag Studies in the Automotive Field

Leading automotive manufacturers in recent years have directed considerable research toward determining how the detailed shapes of various areas of a vehicle (automobile, truck, bus, etc.) affect aerodynamic drag, with particular emphasis on improving fuel economy as well as performance in crosswinds at high speeds. For example, studies have included a combination of aesthetic considerations with aerodynamic performance. Such studies become varied and complex because of the several accepted basic auto designs (fastback, hatchback, notchback rear ends, for example). Computer simulation is playing an increasing role in such research because predictions of aerodynamic flows by computer provide rapid preliminary evaluation of prototype shapes, prior to large-scale testing, as provided by wind tunnels. Many examples of drag coefficient research could be given.

One manufacturer (spring of 1987) devised an improved method for determining projected vehicle frontal areas and thus improved drag coefficient values. See Fig. 12. Examples from bus and truck manufacturers are given in Fig. 13. The application of color computer graphics is shown in Fig. 14.

Theory of Aerodynamic Compressibility

The "incompressible fluid" theory of classical hydrodynamics has proved useful for the estimation of aerodynamic parameters, and when applied to problems of low-speed flight has yielded sufficiently accu-

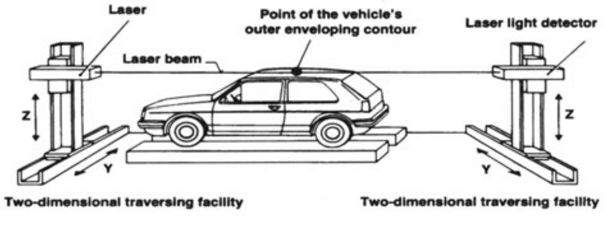
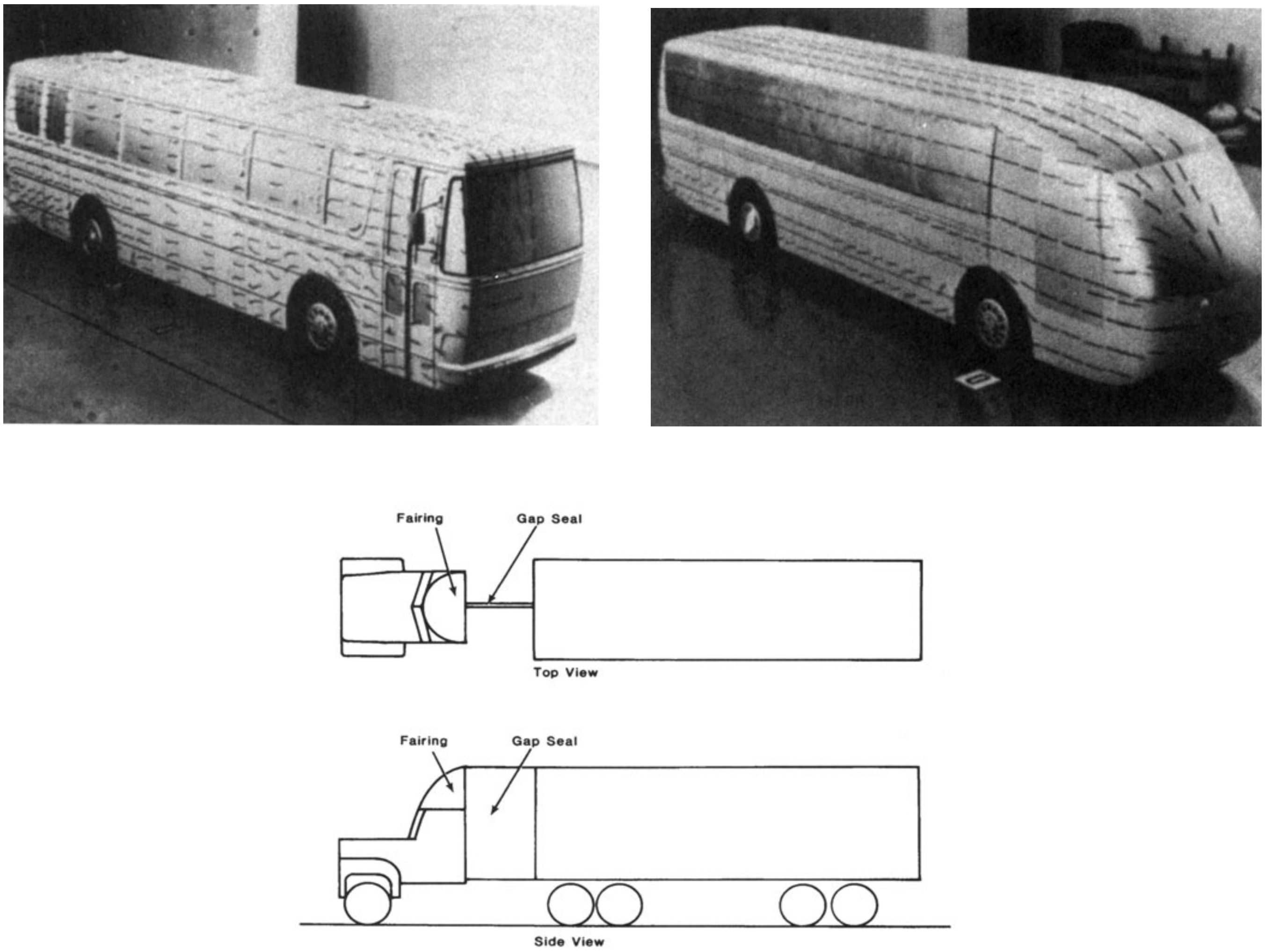


Fig. 12. In determining the drag coefficient of a vehicle, accurate measurements of the vehicle's contour must be made. The drag coefficient of a body in a flowing inviscous fluid equals the ratio of twice the force on the body in the direction of flow to the product of the fluid's density, the square of the flow velocity, and the body's cross-sectional area. Although there are other methods for determining contour, such as shadow imaging; in contour *scanning*, the vehicle's outer enveloping contour can be measured point-to-point by using an optical projector guided perpendicular to the vehicle's longitudinal axis by a numerically controlled traversing facility. When the line of sight of the measuring device strikes the vehicle's contour, an appropriate signal is generated to trigger the coordinate recorder. In system shown here, a helium-neon laser is used as light source. (*Automotive Engineering*.)



(c)

Fig. 13. Tractor-trailers and buses generally are large bluff bodies which do not have ideal aerodynamic shape. Because of the high replacement cost of these vehicles, various researchers have investigated the effects of "add-on" devices on the reduction of aerodynamic drag. One study on buses has shown that drag can be reduced as much as 27%, with fuel consumption reduction of 12%. A traditional bus front-end is shown in (a); an optimized, aerodynamically attractive front-end is shown in (b). This study was made using scaled-down models (50%) and testing in a wind tunnel.

In a study of tractor-trailers, it was concluded that conventional tractors do not have inherently lower drag coefficients than those with cab-over-engine tractors. It has been found that it is the streamlining of the tractor and the interference which it produces with the flow about the trailer that is important, not the type of tractor. As shown by (c), a streamlined cab-roof fairing and the gap seal were evaluated as to their effect on the drag distribution of the tractor-trailer combination. The fairing guides air around and over the trailer so that it does not impact on the front of the trailer and separate at its forward edges. The gap seal is a flat plate mounted vertically along the centerline of the vehicle between the tractor and trailer. It improves the horizontal flow through the gap due to crosswinds and hence reduces separation on the front of the trailer due to this flow. When used in conjunction with the fairing, it also acts to stabilize the wake of the fairing in gusty winds and thus acts to maintain the fairing's effectiveness. The fairing increased the drag on the tractor, but decreased the drag on the trailer by a greater amount, so that the total drag on the vehicle was reduced. The drag on the trailer became zero at 0° yaw due to the low pressure in the gap caused by entrainment. The fraction of the drag due to the tractor with the fairing in place ranged from 100% at 0° yaw to 75% at 20° yaw, including the drag on the fairing.

rate results. It has been found, however, that the flow pattern about a body moving through the air at high speeds is affected to a large degree by changes in density resulting from compression or expansion of the fluid. Consequently, aerodynamic coefficients based on an incompressible-flow theory are in considerable error when applied to airplanes moving at high speeds. An understanding of compressible flows is, therefore, of the utmost importance to the designer of high-speed aircraft.

In the study of compressibility phenomena as applied to aerodynamic problems, airplane speeds are classified according to their relation to the speed of sound in air. The speed of sound is taken as a reference velocity because it is a function of fluid elasticity. As applied to compressible flows this means that the amount of pressure necessary to cause a given change in density in any fluid is proportional to the speed of sound in the fluid. Since the pressure is proportional to the square of the velocity, the velocity which a body may attain before appreciable density changes occur is also proportional to the velocity of sound in the fluid. It is apparent, therefore, that the flow pattern about a body will be altered by density changes to a degree dependent upon the ratio of the velocity of the body to the velocity of sound. This ratio is known as the *Mach number* and is taken as an index of the effects of compressibility on the flow pattern. A curve showing the variation of the speed of sound with altitude is presented in Fig. 15. The local velocity of sound is directly proportional to the square root of the absolute temperature. Therefore, the variation of the speed of sound with altitude is essentially due to the variation in the temperature of the air.

In order to show quantitatively the changes in air density associated with increasing Mach number the following table has been prepared. This table is based on an adiabatic flow through a converging nozzle. M is the stream Mach number and ρ/ρ_s is the ratio of the corresponding air density to the density at static conditions.

M	ρ/ρ_s
0.2	0.981
0.4	0.925
0.6	0.841
0.8	0.742
1.0	0.635

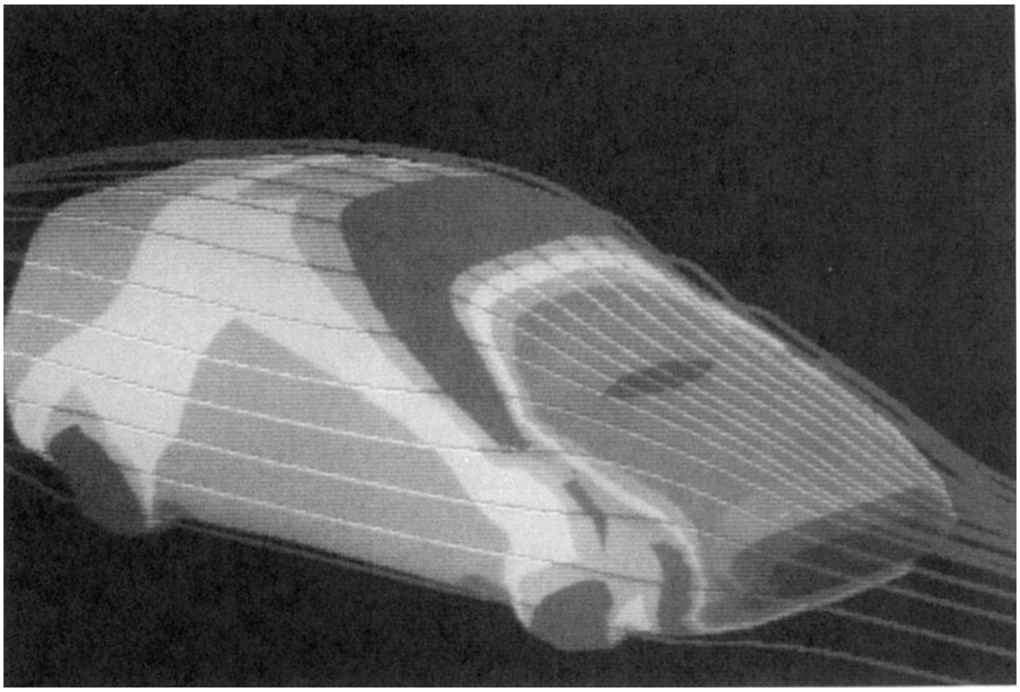


Fig. 14. Facsimile of color computer graphics image showing contours of computed surfaces and streamlines in the flow of an early evaluation of the aerodynamic characteristics of a proposed auto design. (*General Motors Corp.*)

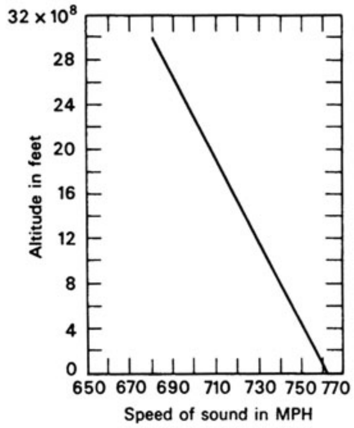


Fig. 15. Speed of sound as a function of altitude. Sea-level temperature, 60°F (15.6°C).

The preceding discussion fully accounts for the effects of compressibility only at Mach numbers less than one, or at subsonic speeds. As air speeds approach and attain the velocity of sound, radical changes occur in the flow pattern which do not result entirely from changes in air density. The flow pattern in a perfect incompressible fluid is instantaneously influenced at all points by pressure changes occurring at any point in the flow field. A consideration of the theory of elasticity as applied to fluids, however, indicates that the effects of small pressure changes in a real fluid are transmitted throughout the fluid in the form of waves which travel at the speed of sound. It may be seen, then, that the effects of a pressure change which occurs behind the critical point at which the speed of sound has been reached cannot influence the flow field ahead of the point.

Since at the critical point the forward motion of the pressure waves is completely arrested by an airstream velocity equal to the velocity of wave propagation, a wave front is formed at the critical point. This wave front constitutes a sharp discontinuity in the flow with which are associated large increases in pressure, density, and temperature and a decrease in velocity. Such a wave front with its attendant discontinuities is known as a shock wave. The flow field about a body traveling at or near sonic velocities will be radically different from that at low speeds.

Once the velocity has increased beyond the sonic range and attained a sufficiently high supersonic value, the shock wave will be forced downstream and the flow at the original point of shock will be comparatively smooth. However, the flow pattern at supersonic speeds will bear little resemblance to that which obtains at low speeds.

High-speed flight is greatly complicated by the compressibility phenomena which have been described. The following paragraphs give an outline of the more important considerations in the aerodynamics of high-speed flight.

The change in aerodynamic characteristics which results from fluid compressibility is greatest when shock waves form on some part of the airplane such as an airfoil or cowling. The forward velocity of the airplane which corresponds to the formation of shock waves is called the critical speed. The critical airplane speed is always less than the velocity of sound since the local velocity over the airfoil surfaces and other components is in excess of the forward speed of the airplane. The critical Mach number is the ratio of the critical airplane speed to the speed of sound.

The airfoil section usually predominates as the factor which controls the effects of compressibility on the characteristics of high-speed airplanes. The variation in pressure distribution around an airfoil as the Mach number is increased is shown in Fig. 16. Both flight and wind-

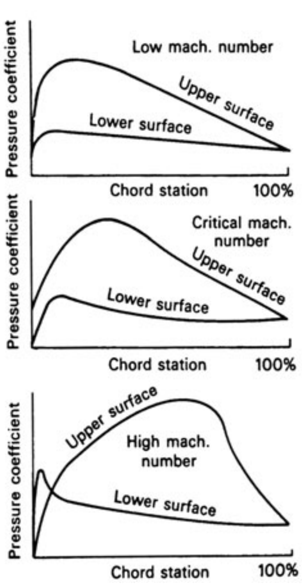


Fig. 16. Pressure distribution of an airfoil section at three different Mach numbers.

tunnel tests have shown that this change in pressure distribution affects the following important airfoil parameters:

(1) The drag coefficient, C_D.
(2) The slope of the lift curve $dC_l/d\alpha$.
(3) The pitching-moment coefficient, C_M.
(4) The maximum lift coefficient, $C_{l\max}$.

The drag coefficient suffers most, its value increasing tremendously as the critical Mach number is reached. The large increase in drag results not only from energy loss in the shock wave but also from the large positive pressure gradient existing across the shock. Such a pressure gradient causes boundary-layer separation which results in a wide, turbulent wake with its attendant-form drag. A curve is presented in Fig. 17 which shows qualitatively the variation of airfoil-drag coefficient with Mach number. In terms of airplane performance, the increase in drag which occurs at the critical speed indicates that a tremendous amount of power would be required for an airplane to fly through the sonic range of speeds and reach speeds above the speed of sound.

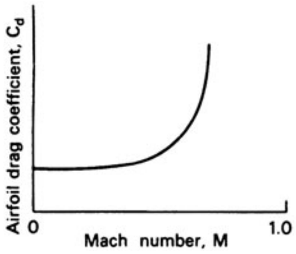


Fig. 17. Variation of airfoil drag coefficient with Mach number.

A loss in maximum lift coefficient occurs as the critical speed is approached. However, airplanes do not operate at high lift coefficients when traveling at high speeds. The changes in lift-curve slope and pitching-moment coefficient which occur with increasing Mach numbers are shown in Figs. 18 and 19. The variation of lift-curve slope with Mach number was calculated from Glauert's approximately correct theoretical relation which states that the change in lift-curve slope is proportional to $1/\sqrt{1 - M^2}$, where M is the free-stream Mach number.

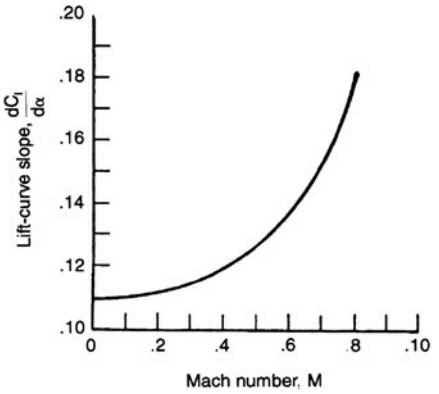


Fig. 18. Change of lift-curve slope with increasing Mach number.

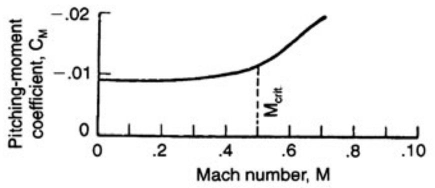


Fig. 19. Variation of pitching-moment coefficient with increasing Mach number.

The changes in pitching-moment coefficient and lift-curve slope indicate a considerable variation in the external forces acting on the lifting surfaces of airplanes traveling at high speeds. Unless the changes in C_M and $dC_l/d\alpha$ are taken into consideration in the design of high-speed airplanes, complete loss of control and stability may result at high speeds. The large turbulent wake discussed in connection with the drag of bodies at shock speeds also has an adverse effect on the control and stability. The angle of downwash and thus the trim of the airplane will change instantly when the wake widens and the plane may be thrown completely out of control. The large turbulent wake may also cause serious control-surface buffeting.

Knowledge of compressible flows is steadily increasing. The most fruitful methods of obtaining information about compressibility phenomena have been found in the field of wind-tunnel testing and through careful instrumentation of rocket-propelled test vehicles for which the measured data are telemetered to the ground.

Theory of Aerodynamic Circulation

A mass of air in rotary motion is said to be in circulatory flow if its velocities at various radii from the center of rotation are of the proper magnitude to induce radial equilibrium of the circulating mass. Consideration of the requirements for equilibrium consists in balancing centrifugal forces against static pressures derived from Bernoulli's theorem and results in the specification that velocity of a particle must be inversely proportional to its radius from the center of rotation. Note that this is not like the motion of portions of a wheel, whose velocities are proportional to their radii. A simple case of circulation could be visualized as the flow pattern (concentric circles) induced by a rough cylinder rotating rapidly in still air.

If air is in circulation and no object such as a cylinder occupies the central core, the velocity at the center of rotation reaches a theoretical value of infinity. This is impossible and the center of circulatory flow must be occupied by a small core of air in simple rotary motion. Air in this condition is described as a free vortex. Exceptionally high velocities may exist at the edge of the rotary core, and in vortices of high circulatory strength such as tornadoes the atmospheric energy may reach the destructive proportions.

The strength of aerodynamic circulation (usually called the *circulation* and designated by symbol Γ) is the line integral of the tangential component of the velocity along any closed line encircling the vortex center, or some object.

$$\Gamma = \oint v dl$$

It has been proved that Γ has the same value for any closed path through the flow around the object (such as an airfoil, cylinder, etc.) and this fact is of value in various phases of aerodynamic analysis. A simple case of a flow pattern of concentric circles illustrates the constancy of Γ. The velocity at radius r is K/r and

$$\Gamma = \oint \frac{K}{r} dl = \oint \frac{K}{r} r d\theta = 2\pi K$$

Therefore Γ is not a function of r; its magnitude is constant throughout the flow pattern.

The generally accepted demonstration of the origin of lift on an airfoil employs a vortex circulation imposed on a rectilinear velocity field to produce the typical stream flow pattern, after which the aerodynamic forces are analyzed as originating from the impulse required to alter the momentum of the airstream and the static pressure. Fortunately, it is possible to analyze lift on the premise of frictionless flow. Viscosity effects do not enter until drag is sought. The dependence of lift of an airfoil of infinite span upon circulation Γ, free stream velocity V, and mass density ρ was proved independently by Kutta and Joukowski early in this century. Later Prandtl and others extended the "circulation theory" to cover the lift of finite wings. Joukowski's proof, considerably abbreviated, follows. The inherent tendency of an airfoil to create circulation is replaced by a bound vortex. A section of airfoil of unit span length is taken. The two flows are shown covering the same region in Fig. 20. Of course this figure does not represent the actual stream flow pattern since the latter would follow the vectorial combination of these two flows. Consider the region enclosed by the imaginary cylindrical

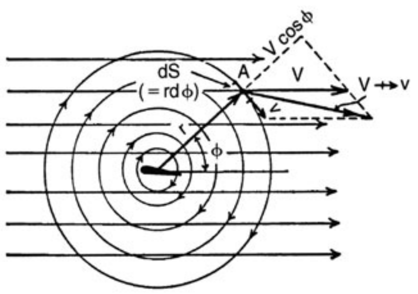


Fig. 20. Circulatory superimposed on rectilinear motion.

surface at radius r from the vortex center. Let r be large enough to have circulatory velocity v small compared to rectilinear velocity V and integrate for vectorial change in momentum in this cylinder and vertical component of static pressure acting against it.

The velocity of circulation, v, is $\Gamma/(2\pi r)$ and its combination with V gives the direction and magnitude of the stream line at point A. The mass of air leaving the cylindrical region at A is $\rho ds\ (V \cos \phi)$ and its vertical component of velocity is $v \cos \phi$. The net vertical momentum in the cylindrical mass of air, due to inflow and outflow, is

$$\oint (\rho r d\phi)(V \cos \phi)\left(\frac{\Gamma}{2\pi r} \cos \phi\right)$$

which simplifies to

$$\frac{\rho \Gamma V}{2\pi} \oint \cos \phi d\phi$$

whence, by integration, the vertical impulse force is $(\rho\Gamma V)/2$. Since the change of momentum is downward the impulse force is upward, i.e., it is the lift. An integration of the horizontal component of momentum yields zero net momentum.

Proceeding next to evaluate the static pressure, note that Bernoulli's theorem applied at point A is covered by the following statement. Let p be the static pressure at A, and p_0 the free stream static pressure; then

$$p + \frac{\rho}{2}(V \mathbin{+\!\!\!\rightarrow} v)^2 = p_0 + \frac{\rho}{2} V^2$$

After expanding $(V \mathbin{+\!\!\!\rightarrow} v)^2$ and dropping all v^2 terms (since A was chosen to make v small relative to V),

$$p = p_0 - \rho V v \sin \phi$$

Since the second term on the right-hand side of this equation represents the variation from free stream pressure at point A, the lift (if any) will be the line integral of its vertical component; i.e.,

$$\begin{aligned} \text{vertical pressure component} &= \oint (\rho V v \sin \phi)(\sin \phi)(r d\phi) \\ &= \rho \frac{V\Gamma}{2\pi} \oint \sin^2 \phi d\phi \\ &= \rho \frac{V\Gamma}{2} \end{aligned}$$

The horizontal integral yields zero net pressure. The total lift per unit of span is the sum of that originating from momentum and that originating from pressure. From the preceding demonstration it is seen that each contributes equally to the total lift, which becomes $\rho V\Gamma$.

Theory and experiment show that Γ is always sufficient to cause the divided flow over the upper and lower surfaces to reunite at the trailing edge without reverse flow.

If the aerodynamic chord of the wing is parallel to the free stream velocity V, the lift is zero and obviously Γ is zero. As the angle of attack is increased Γ increases nearly proportionately. The strength of the circulation for an airfoil of chord c at aerodynamic angle of attack α_a is $\frac{1}{2} a V c \alpha_a$, where a is the proportionating factor. Since lift, $L, = \rho V\Gamma b$ (b = span), it also equals

$$a\alpha_a \frac{\rho V^2}{2} (bc)$$

Call $a\alpha_a$ the coefficient of lift, bc the area S, and $(\rho V^2)/2$ the dynamic pressure q; then the lift equation becomes

$$L = C_L q S$$

If the wing has a finite span (i.e., definite tips), the bound vortex, assumed for the purpose of accounting for lift, is found to extend from the tips downstream as a free vortex. The effect of these tip vortices is to modify the air reaction since a downwash, w, is produced at the lifting line (see Fig. 21). The effect of downwash is to change the lift direction from L to L' since the relative wind, V_r, is now inclined. The angle ϵ is always small, so L' may be considered of the same magnitude as L. However, L still remains the significant lift since the free stream velocity is V_0. Consequently, a drag D_i is introduced, although the original assumption of zero air viscosity remains in order.

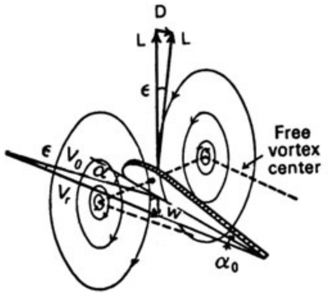


Fig. 21. Perspective of a section of airfoil subject to downwash from the free vortices at the tips.

This *induced drag* is a feature of the motion of ideal, frictionless air over a wing of finite span. The longer the span, the less the average effect of downwash (since in vortex motion $v \sim r^{-1}$) and the less the induced drag, but it will not disappear unless the span extends to infinity. By application of hydrodynamic theory beyond the scope of this article, aerodynamicists have shown that:

(1) Minimum induced drag for a given lift and span will be had if downwash is of constant magnitude over the span.

(2) Constant downwash will result from a bound vortex circulation strength of elliptical spanwise character.

(3) Elliptical spanwise Γ occurs on an untwisted airfoil of elliptical planform.

(4) With elliptical Γ a simple relation connects induced drag and aspect ratio.

$$D_i = L^2/\pi \mathrm{AR}$$

where AR represents the aspect ratio.

This relation holds for the coefficient, too.

$$C_{Di} = C^2{}_L/\pi A\mathrm{R}$$

As

$$\frac{D_i}{L} = \frac{C_{Di}}{C_L} = \frac{w}{V_0} \text{ (nearly)}$$

$$\alpha = \alpha_0 + C_L/\pi A\mathrm{R}$$

It appears that a finite wing possesses an induced drag of C_{Di} not possessed by an infinite wing in ideal flow, also an induced angle of attack, requiring the angle of attack α of the finite span to be more than α_0 of an infinite span for the same lift per unit of span. The differences of induced drags of two wings of the same airfoil profile at the same lift coefficient but with different planforms follow naturally from the above:

$$C_{Di1} - C_{Di2} = \frac{C_L^2}{\pi}\left(\frac{1}{A\mathrm{R}_1} - \frac{1}{AR_2}\right)$$

It is commonly assumed that the difference between the total drag coefficients of two wings of the same airfoil profile but different aspect ratios is the difference in induced drag. Likewise, the difference in necessary angle of attack to the free stream velocity for the same C_L is:

$$\alpha_1 - \alpha_2 = \frac{C_L}{\pi}\left(\frac{1}{A\mathrm{R}_1} - \frac{1}{AR_2}\right)$$

While these induced effects are premised on elliptical planform for the wing, they hold very well for rectilinear wings with elliptical tips and those with moderate taper ratio.

The lateral or spanwide distribution of lift on a wing depends on the planform and the distribution of downwash. Two cases are illustrated—an elliptical wing and a rectangular one. (See Fig. 22.)

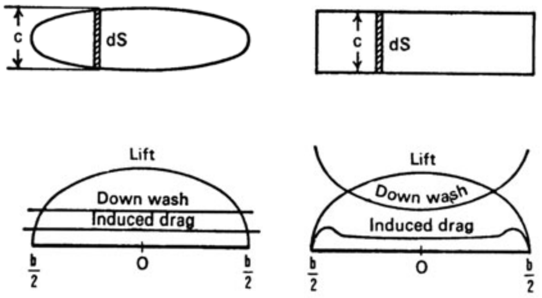


Fig. 22. Spanwise variations.

(1) The effect of downwash is to reduce the effective angle of attack and hence the lift coefficient. The lift of any lateral element *dS* is proportional to the product of C_L and *c*; hence the lift distribution of an elliptical wing is elliptical, while its induced drag, mirroring the downwash, is constant.

(2) The downwash strength increases near the tips of a rectangular wing. The lift coefficient accordingly declines and lift falls off at the tip even though chord is maintained. Induced drag is strong at the tips.

Supersonic Aerodynamics

During World War II, certain fighter aircraft attained speeds which produced, on critical points of the airplane, local flow velocities of Mach 1.0 and higher, even though the airplane speed was less than that of sound. Since that time, much greater speeds have been attained. This performance introduced a new dimension to the science of aerodynamics that is better addressed in a separate article devoted exclusively to that topic. See **Supersonic Aerodynamics.**

For further coverage of aerostatics, see articles on **Balloon;** and **Dirigibles and Airships.**

Additional Reading

Beardsley, T.: "Earning Its Wings: Hypersonic Flight," *Sci. Amer.*, 17 (June 1988).
Culick, F. E. C.: "The Origins of the First Powered Man-carrying Airplane," *Sci. Amer.*, 86 (July 1979).
Dole, C. E.: "Flight Theory and Aerodynamics," Wiley, New York, 1991.
Goehler, D. D.: "Aerospace Industries Materials," *Advanced Materials and Processes*, 77 (January 1990).
Hull, D.: "Energy Absorbing Composite Structures," *University of Wales Review*, 23 (Spring 1988).
Jameson, A.: "Computational Aerodynamics for Aircraft Design," *Science*, 361 (July 28, 1989).
Kuethe, A. M., and C-Y Chow: "Foundations of Aerodynamics," Wiley, New York, 1985.
Lukasak, D. A., and R. M. Hart: "Strong Aluminum Alloy Shaves Airframe Weight," *Advanced Materials and Processes*, 46 (October 1991).
March, A.: "The Future of the U.S. Aircraft Industry," *Technology Review (MIT)*, 26 (January 1991).
Staff: "Frontal Areas Measured More Precisely (Drag Coefficient)," *Automotive Eng.*, 76 (March 1987).
Staff: "Advanced Materials in Aerospace Applications," ASM International Materials Park, Ohio, 1990.
Stephens, J. R.: "Composites Boost 21st Century Aircraft Engines," *Advanced Materials and Processes*, 35 (April 1990).
Stix, G.: "Plane Geometry: Boeing Uses CAD to Design 130,000 Parts for Its New 777," *Sci. Amer.*, 110 (March 1991).
Stix, G.: "Smaller World. The Draper Prize Recognizes the Father of the Jet Age," *Sci. Amer.*, 57 (December 1991).
Strauss, S.: "Like A Bird," *Technology Review*, 8 (August 1989).
Suter, A. M.: "Noise Wars," *Technology Review (MIT)*, 47 (November 1989).
Wootton, R. J.: "The Mechanical Design of Insect Wings," *Sci. Amer.*, 114 (November 1990).

AEROELASTICITY. The study of both the static and dynamic effects of aerodynamic forces on elastic bodies. The swaying of bridges, trees, smokestacks, and buildings are examples of the interplay of aerodynamic forces, inertia forces, and elastic properties of the structures. The flutter of flags, aircraft wings, and sails are more examples. For aircraft design, aeroelasticity is a most important study.

AEROGEL. A colloidal solution of a gaseous phase in a solid phase, obtained usually by replacement of the liquid in the dispersed phase by air or gas. Contrast with **Aerosol.**

AEROLITE. A general term for meteorites that are richer in the basic silicates than in nickel and iron. See also **Meteoroids and Meteorites.**

AEROLOGY. See **Meteorology.**

AEROSOL. A colloidal system in which a gas, frequently air, is the continuous medium, and particles of solids or liquid are dispersed in it. Aerosol thus is a common term used in connection with air pollution control. Studies of the particle size distribution of atmospheric aerosols have shown a multimodal character, usually with a bimodal mass, volume, or surface area distribution and frequently trimodal surface area distribution near sources of fresh combustion aerosols. The coarse mode (2 micrometers and greater) is formed by relatively large particles generated mechanically or by evaporation of liquid from droplets containing dissolved substances. The nuclei mode (0.03 micrometer and smaller) is formed by condensation of vapors from high-temperature processes, or by gaseous reaction products. The intermediate or accumulation mode (from 0.1 to 1.0 micrometer) is formed by coagulation of nuclei. Study of the behavior of the particles in each mode has led to the belief that the particles tend to form a stable aerosol having a size distribution ranging from about 0.1 to 1.0 micrometer. The larger, settleable particles (in excess of 1.0 micrometer in size) fall out, whereas the very fine particles (smaller than 0.1 micrometer) tend to agglomerate to form larger particles which remain suspended. The nuclei mode tends to be highly transient and is concentration limited by coagulation with both other nuclei and also particles in the accumulation mode. It further appears that additional growth of particle size from the accumulation mode to the coarse mode is limited to 5% or less (by mass). Thus, the particulate content of a source emission and the ambient air can be viewed as composed of two portions, i.e., settleable and suspended.

Both settleable and suspended atmospheric particulates have deleterious effects upon the environment. The settleable particles can affect health if assimilated and also can cause adverse effects on materials, crops, and vegetation. Further, such particles settle out in streams and upon land where soluble substances, sometimes including hazardous materials, are dissolved out of the particles and thus become pollutants of soils and surface and ground waters. Suspended atmospheric particulate matter has undesirable effects on visibility and, if continuous and of sufficient concentration, possible modifying effects on the climate. Importantly, it is particles within a size range from 2 to 5 micrometers and smaller that are considered most harmful to health because particles of this size tend to penetrate the body's defense mechanisms and reach most deeply into the lungs.

The term aerosol is also applied to a form of packaging in which a gas under pressure, or a liquefied gas which has a pressure greater than atmospheric pressure at ordinary temperatures, is used to spray a liquid. The result of the spraying process is to produce a mist of small liquid droplets in air, although not necessarily a stable colloidal system. Numerous products, such as paints, clear plastic solutions, fire-extinguishing compounds, insecticides, and waxes and cleaners, are packaged in this fashion for convenience. Food products, such as topping and whipped cream, also are packaged in aerosol cans.

For a number of years, chlorofluorocarbons were the most popular source of pressure for these cans. Because of concern in recent years over the reactions of chlorofluorocarbons in the upper atmosphere of the earth that appear to be leading to a deterioration of the ozone layer, some countries have banned their use in aerosol cans. Manufacturers have turned to other gases or to conveniently operated hand pumps. See **Ozone.**

See also **Colloid System;** and **Pollution (Air).**

AESTIVATION. Summer dormancy, the antithesis of the more familiar hibernation. As an example, the African lungfish is known to burrow down into the mud when the water begins to get low in the rivers and lakes where it lives. Here, even though the mud dries into a hard cake, the lungfish survives in a state of aestivation. Some have even been cut out along with the surrounding dirt and shipped abroad. When the mud again becomes moist with water, the lungfish becomes active and resumes normal activity.

AFFINE TENSOR AND FREE VECTOR. A quantity that behaves like a tensor under a linear (affine) coordinate transformation, but not under a general coordinate transformation is called an affine tensor. From an affine tensor it is possible to construct a *free vector*, that is, a vector not related to a given point (non-localized vector).

AFLATOXIN. See **Yeasts and Molds.**

AFRICAN TRYPANOSOMIASIS. An infection by the flagellate blood protozoan parasite *Trypanosoma brucei* which causes a fatal neurological disease whose final stage in humans is *sleeping sickness*. There are two epidemiological and clinical variants of the parasite: *T. rhodesiense* and *T. gambiense*, which are carried by various species of tsetse fly (*Glossinia*) in Africa between latitudes 15°N and 20°S. About 50 million people are at direct risk of contracting the disease, some 20,000 new cases being reported every year and thousands more going unreported. Even more important than the direct threat to humans is the fact that domestic livestock are susceptible to trypanosomiasis. Between them, the trypanosome and the tsetse fly make some four million square miles (10.4 mil sq km) of Africa uninhabitable for most breeds of dairy and beef cattle. However, dwarf, indigenous cattle and goats live and survive within the tsetse zone. Having little or no access to meat and dairy products, most of the human population is malnourished and susceptible to other diseases.

The Gambian form of trypanosomiasis, found in West and Central Africa, is without a known natural reservoir and is a milder, endemic disease associated with fly breeding near streams. The Rhodesian sleeping sickness of East and Central Africa is clinically more severe and is carried by tsetse flies that breed in savannas and brushlands.

When the trypanosome is ingested by the tsetse fly, it lodges in the midgut and undergoes a series of biochemical and structural changes. When the fly bites a mammal, metacyclic trypanosomes are introduced into the host's blood stream, where they rapidly differentiate to a form which can proliferate. In the human blood, cerebral fluid, lymph nodes, and other organs, the trypanosome appears as a motile, pleomorphic organism 15–30 micrometers long with a flagellum and undulating membrane.

In humans, the disease is characterized by fever, insomnia, lymph node enlargement, local edema and rash. After about six months of infection, there may be invasion of the central nervous system producing somnolence, confusion, wasting, coma and death. When untreated, the disease is fatal within one year.

Trypanosomiasis defies vaccination because the human immune response cannot protect against the infection. Typically, a fly will inject trypanosomes into the host's blood stream and the host will manufacture an enormous number of antibodies, enough to destroy at least 99% of the trypanosomes within a week or so, but the remaining few trypanosomes will elude the human defense by changing the antigen which constitutes their surface coat. These are effected by what is known as the variable surface glycoprotein (VSG). By the time the immune system has made new antibodies to bind to these new antigens, some of the trypanosomes have shed their coats again and replaced with yet another antigenetically distinct one. The host's overworked immune system is unable to cope with the infection and so the parasite proliferates.

Drugs currently available for treating trypanosomiasis—suramin, pentamidine, or melarsoprol—are extremely toxic and also cannot prevent reinfection. New forms of chemotherapy are being sought. The trypanosomes cannot survive in the mammalian host without their protective surface coats. Therefore, drugs which interfere with the phosphoglyceride anchoring the VSG in the cell membrane, or which activate the enzyme releasing the VSG, may be therapeutic agents. Mammalian messenger RNAs do not have the specific sequence of 35 neucleotides required for the VSGs and a drug which interferes with the synthesis of mRNA may therefore disable the parasite. One of the latest additions to the drugs for fighting the infection is alpha-difluoromethyl ornithine (DFMO), which is a polyamine synthesis inhibitor developed as an anti-tumor agent. The drug has been used dramatically in terminally ill patients and may be more potent against trypanosomiasis if combined with drugs, such as suramin.

Other approaches to elimination of the disease have involved eradication of the tsetse fly by spraying areas with insecticide or by distribution of sterile male flies. These methods, however, are expensive and have not been effective over extremely large territory as previously mentioned.

Additional Reading

Clarkson, A. B., et al.: *Amer. J. Trop. Med. and Hyg.*, **33**, 1073 (1984).

Clarkson, A. B., et al.: *Proceedings National Academy of Sciences (USA)*, **80**, 5729 (1983).

Clarkson, A. B., et al.: *Science*, **227**, 118–119 (1985).

Donaldson, J. E., and M. J. Turner: "How the Trypanosome Changes Its Coat," *Sci. Amer.*, **252**(2), 44–51 (1985).

Kolata, G.: "Scrutinizing Sleeping Sickness," *Science*, **226**, 956–959 (1984).

Pearson, M., Nelson, R. J., and N. Agabia: *Immunology Today*, **5**(2), 43 (1984).

Strickland, G. T.: "Hunter's Tropical Medicine," 7th Edition, W. B. Saunders, Philadelphia, Pennsylvania, 1991.

Warren, K. S., and A. A. F. Mahmoud, Eds.: "Tropical and Geographic Medicine," McGraw-Hill, New York, 1984.

Ann C. Vickery, Ph.D., Assoc. Prof., College of Public Health,
University of South Florida, Tampa, Florida.

AFTERBIRTH. The membranes and placenta expelled from the uterus a short time after the birth of the child. See **Embryo.**

AFTER-IMAGE. The image "seen" after a portion of the retina has been fatigued by continued fixed stimuli. For instance, if a person stares fixedly at a black cross-mark on a sheet of white paper for a minute or two and then suddenly looks at a blank wall, he will "see" a white cross-mark on the wall. If the image he stares at is colored, he will usually see an image of the complementary color on the wall. A green cross on the paper, for instance, will result in the appearance of a red cross on the wall. This is because the green-sensitive cones of the retina have become fatigued but the red-sensitive cones have not. The red-sensitive cones pick up the red light reflected from the wall, but the green-sensitive cones do not pick up the green light in a proper proportion.

AFTERSHOCK. See **Earth Tectonics and Earthquakes.**

AGAMIDS. Of the class *Reptilia,* subclass *Lepidosauria,* order *Squamata* (scaly reptiles), suborder *Sauria* (lizards), infraorder *Gekkota,* and family *Agamidae,* according to the classification of Grzimek (1972), the agamids are the counterpart in the New World to the iguanas of the Old World. See also **Iguanas**.

In most cases, these animals are strongly built, with long legs, a rather large head, and a long tail. Usually, the tail cannot be discarded and regenerated, as is the case with many lizards. The body and tail are often covered by sturdy scales, almost always with a keel or ridge and sometimes with spines; the scales on the head are small and are not arranged symmetrically. The tongue is short and fleshy. Vision is acute, playing a greater role than the sense of smell. There are distinct eyelids covered with scales, and desert animals often have scales along the edges like eyelashes to protect the eyes against blowing sand. The auditory organ usually has an external eardrum. The diet consists mainly of insects; the adults of larger forms are often herbivores or omnivores. There are 34 genera, comprising altogether some 300 species.

Many species of these decidedly diurnal animals, particularly the larger forms, live in very dry, even desert-like regions. However, in southeastern Asia, countless agamids live on the ground, in trees, and even in the water of the tropical rain forest.

In general, agamids behave in such a way as to prevent excessive fluctuations in body temperature. Many desert dwellers sun themselves in the morning by positioning themselves perpendicular to the rays of the sun, so that they receive the most radiant energy. See Fig. 1. At the warmest time of day, in contrast, they make their bodies narrow and orient toward the sun, often with the belly raised as far as possible off the hot sand. In such desert animals, the scales on the belly are usually light in color, in order to absorb the least radiation from the sand or rock below. Other species live in crevices in large rocks or in a cave, and they retire to these shelters for a while each day to avoid the midday heat. The ground dwellers, in contrast to the aboreal agamids, are not laterally flattened, but rather are flat on the back and belly. Their coloring is also related to heat regulation; among the desert forms, one finds very dark lizards, colored so that they may lose excess heat by radiation more quickly when they are in the shade.

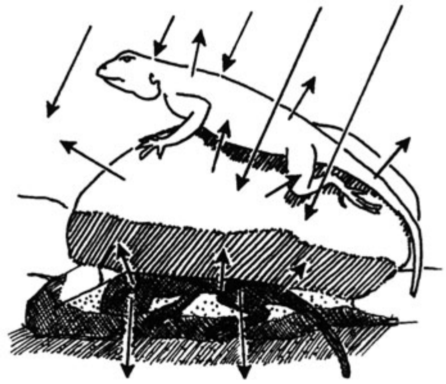


Fig. 1. The common agama controls its temperature in a sensible manner—warming up to its preferred temperature in the morning and positioning its body perpendicularly to the solar radiation, and cooling off by removing excess heat in the shade of a rock. (*After H. Schifter.*)

Agamids usually can change color as strikingly as chameleons. The colors of the head, in particular, reflect several different moods, including fear, eagerness to attack, and sexual excitation. Signals important in social behavior are the characteristic bobbing of the head and the less common waving of the forelegs. The males, in particular, bob their heads at regular intervals to display to a female and to intimidate possible intruders. See Fig. 2.

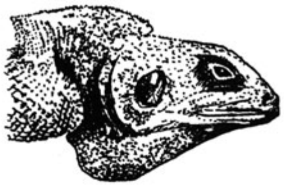


Fig. 2. Head of a male common agama showing the threat display. (*After H. Schifter.*)

Phylogenetic development of the agamids centered in the islands of Indonesia and on the southeastern Asian mainland. From there, certain ground-dwelling genera proceeded to colonize the Near East, Africa, and Australia. In the west it was primarily the genus *Agama*, which settled throughout Africa and the Near East. In the deserts of northern Africa and the Middle East, the spiny-tailed lizards are also found, and *Phrynocephalus* lives in western and central Asia. The range of the mountain-dwelling genus *Japalura* extends as far as northern China. The well-known moloch of Australia (genus *Moloch*) differs so radically from all other agamids that it probably has a long history of evolving in isolation. Later on, other agamids migrated to Australia, including, for example, the bearded lizard, the frilled lizard, the semi-aquatic *Physignathus*, and *Gonocephalus.* This last genus, which also inhabits southeastern Asia, is to be found on a number of Polynesian islands as well.

Some of the approximately sixty species of the agamids include:

1. The *Common Agama.*
2. The *Black-Necked Agama*, whose coloration is variable, but is predominantly green during courtship. His habitat is the savannas of southern and eastern Africa.
3. *Kirk's Rock Agama* has a low crest on the back, and the head of the male is often a bright orange. It dwells in the rocky desert areas, ranging from Zaire to Tanzania.
4. *Bibron's Agama* has a basically yellow-brown coloration. It lives in the rocky deserts of Morocco, Algeria, and Tunisia.
5. *Desert Agama* has rather inconspicuous coloring, except for displays of magnificent blue shading in the male during courtship. The head is short and, as a desert dweller, has well-developed eyelids. It feeds primarily on grasshoppers. Distribution is along the northern edge of the Sahara and the Arabian desert, from Morocco to Iraq.
6. The *Hardun* has legs and tail that are covered with scales bearing pronounced keels. Although predominantly insectivores, the hardun will also eat plants. It is found chiefly in rocky areas, often close to human settlements. Originally distributed from Arabia to western Turkey and on numerous islands of the Aegean Sea, more recently the hardun has been inadvertently transported to northern Greece, Corfu, and the vicinity of Alexandria (Egypt).
7. *Agama agilis* has eyelash-like scales that are similar to the conical lids of chameleons. The eyes are independently movable and thus can fixate different objects simultaneously. This agama inhabits southwestern Asia, from Arabia to Pakistan.
8. *Caucasian Agama* are found in groups of two to twenty-five animals that live and hunt together within a fairly well-defined territory, commonly at altitudes of up to 2700 meters, from Caucasia to Pakistan. In its homeland it is found everywhere in close association with humans, having succeeded in turning the continual destruction of the tropical African forests to its own advantage.
9. *Spiny-tailed lizards* live in rocky desert regions from Senegal to Egypt. They are quite variable in coloration, usually blackish with spots and cross bands which may be yellow, orange, red, and, less frequently, green. They are quiet animals and are easily supported in captivity. The variable color of the skin enables these lizards to carefully control heat radiation which they receive. They are usually quite dark in the morning (to absorb radiation) and of a light color in the afternoon (to absorb less radiation).
10. *Toad-Headed Agamids* are among the smallest of the agamids. They range from Caucasia and Arabia to western China.
11. The *Moloch* of Australia is remarkable for its frightening appearance, with its body and tail covered with large spiny scales. Although of a dangerous appearance, the animal is quite harmless. It consumes ants at an exceedingly rapid rate, sometimes consuming nearly 2000 for one meal. The animal accommodates to arid conditions by using its highly specialized skin structure. The skin is equipped with threadlike canals between the scales. These canals suck up water in seconds, but do not discharge directly through the body wall because, if the skin were permeable, body water would escape. Rather, in a slow manner, the canals discharge into the corners of the mouth, stimulated by the animal's mouth movements.
12. *Frilled Lizard* has no dorsal crest, but does have an enormous "frill" (a fold of skin below either side of the head, supported by a row of cartilaginous extensions). Normally, the frill is kept folded back against the body, but during threat and courtship displays, the mouth is opened and the highly colored frill expands in a dramatic manner.
13. *Flying Dragons* (genus Draco) is unique among all present-day reptiles because of its specially adapted means for moving through the air. These "dragons" have two large, winglike flaps of skin, one on each flank. The last five to seven ribs are long, forming the movable skeleton that supports these "wings," which usually are folded flat along the body, but when the animal spreads them, it can perform long gliding flights from tree to tree. Other features of the body structure appear to have been "stylized" in favor of lightness and flying ability. There are sixteen species of *Draco*. All inhabit the tropical rain forest, often living high in the crowns of trees. Although flights of 60 meters have been observed, normally the distance is about 10 meters. *Draco* is widespread on the Indonesian islands.

The Chamaeleons. Most zoologists regard these animals as descendants of the ancient Agamidae because fossil forms have been found as early as the Cretaceous period. In addition to an ability to change their color, they have numerous distinct peculiarities, including their remarkable eyes, which can be moved independently. Each eye has a scaly lid in the form of a cone, with only a small opening in the middle for the pupil. Also, the tongue can be shot out to an extraordinary length, sometimes equal to the length of the animal. This is a great advantage for capturing prey at a comparatively long distance. Chameleons also are well adapted to arboreal life. The feet are modified and oppose one another much like the thumb and fingers of primates. The tail is a prehensile device, but, unlike many lizards, it cannot be broken off and subsequently replaced. Their diet includes spiders, grasshoppers, and stick insects, although in a terrarium they will accept other forms of insects. Larger chameleon species also consume small lizards and nesting mice. Their requirements for water generally are satisfied by licking the dew from foliage.

Chameleons are egg layers—up to thirty or forty at a time, which are buried in the ground. However, some mountain species are ovoviviparous. Interesting characteristics of chameleons are shown in Fig. 3. The ovoviviparous process of the dwarf chameleon is shown in Fig. 4.

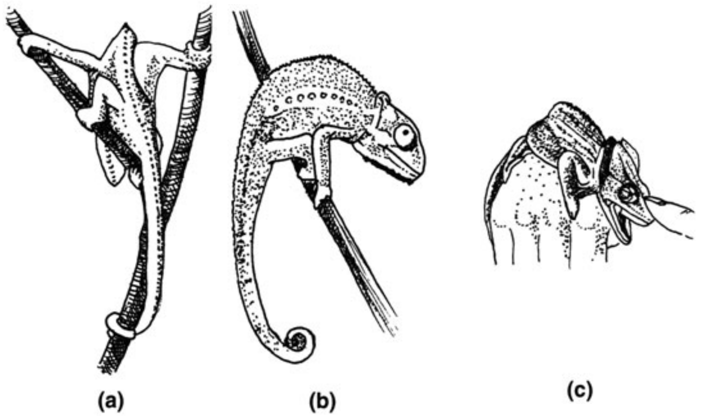


Fig. 3. Some characteristics of the dwarf chameleon (*Chamaeleo pumilus*), whose habitat is southern Africa. (a) To avoid falls, the animal climbs very carefully and anchors its body at five points, including the tail. (b) An old male dwarf chameleon. (c) Aggressive posture of the short-horned chameleon (*C. brevicornis*), whose habitat is Madagascar. Lobes behind the head are erected when irritated by a finger in a terrarium. (*After H. Schifter.*)

Some of the numerous species of chameleons include:

1. The *African Chameleon* may be up to nearly 40 centimeters in length. It ranges from western Africa to Ethiopia and Somalia, where it is subjected to annual climatic changes. Thus, the terrarium keeper must accommodate for such changes.
2. The *common chameleon* is found in tropical and southern Africa. It is aggressive and lives preferentially on low trees and bushes, but also descends to the ground, either to change its location or to dig a hole some 20 centimeters deep, in which it lays thirty to forty eggs. The principal enemy is the boomslang snake, which is seldom deterred by the chameleon's defensive mechanisms, including its changing displays of coloration, expanding dewlap, and hissing noises.
3. The *African Two-Lined Chameleon* is ovoviviparous. The animal is light to dark brown, with stripes along the sides of the body. Ten to twenty-five young are born alive at a time. The babies are only about 4 centimeters long, but grow to their full size of 11 to 16 centimeters in less than one year.
4. The *Dwarf Chameleon* is also ovoviviparous and is found in the cooler mountainous regions of southern Africa. Kastle observed that a dwarf chameleon runs away in a crouched posture, with the tail stretched out to impress and threaten its conspecifics, and that, when threatened, it expands its throat, opens its mouth, and rocks its body from side to side while wagging the head. The wagging motion, like the bobbing of the head in other species, also serves as a preliminary to mating.

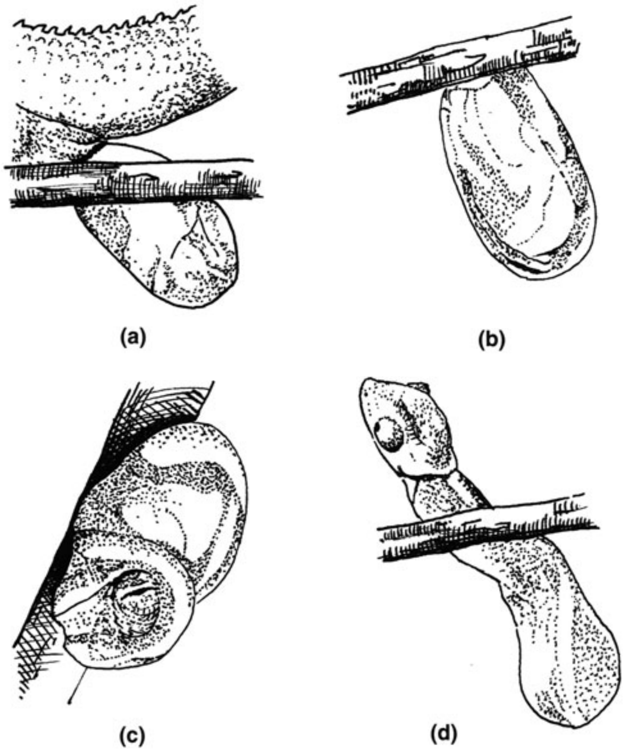


Fig. 4. The ovoviviparous birth of a dwarf chameleon. (a) Egg leaves the cloaca. (b) Egg sticks to a twig. (c) Baby inside egg membrane. (d) Young chameleon twists out of egg, leaving behind the empty shell with remnants of yolk. (*After H. Schifter.*)

5. The *Outstalet's Chameleon* lives on the island of Madagascar and ranges up to 63 centimeters in length. The creature is described as a quick and snappish animal, with a marked requirement for warmth. It principally feeds on large locusts, but can be adapted to a terrarium by feeding it with small mammals and lizards.

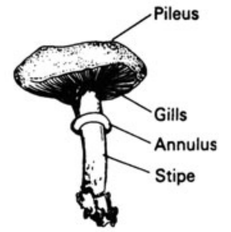
Other chameleons include the *Panther Chameleon,* found on the islands of Reunion and Mauritius; *Parson's Chameleon*, characterized by two processes on the snout; *Meller's Chameleon*, which is the largest chameleon outside of Madagascar; and the *Stumped-Tailed Chameleon* of Madagascar, among others.

See also **Lizards**.

AGAR-AGAR. Now more commonly called "agar." A gelatine-like substance which is prepared from various species of red algae growing in Asiatic waters. The prepared product appears in the form of cakes, coarse granules, long shreds, or in thin sheets. It is used extensively alone or in combination with various nutritive substances, as a medium for culturing bacteria and various fungi. See also **Gums and Mucilages.**

AGARICS (*Agaricaceae; Fungi*). This family of fungi is probably better known than any other, since it contains most of the plants commonly described by the names toadstool and mushroom, which popularly and mistakenly denote poisonous and edible fungi, respectively.

The Agarics are mostly fleshy fungi of that very definite structure, the familiar parasol-like toadstool. This is composed of convex pileus or cap, usually supported on an evident stalk. The underside of the cap shows a series of radiating plates, or gills, which are formed in agaric fungi only, and so serve to separate them from all others. The two sides of the gills are covered with the microscopic spore-bearing bodies called basidia, which are club-shaped or cylindrical cells bearing spores, generally four each. Agarics vary in size from delicate species with a cap a millimeter or so in diameter, supported by a slender threadlike stalk, to massive forms twelve inches (30.5 centime-



Sporophore of a mushroom (*Agaricus campestris*).

ters) in diameter: the larger species form millions of spores. See accompanying figure.

The spores float in the air for considerable distances, and finally come to rest on some solid substance. Should this be favorable for germination, the spore puts out a slender tube, which elongates rapidly and penetrates the substratum, from which it absorbs substances necessary for its continued growth. Gradually this threadlike body, known as the mycelium, spreads through extensive masses of substratum, branching frequently as it does so.

Finally there is accumulated in the mycelium a supply of reserve food sufficient for fruiting: then, if atmospheric conditions, such as moisture and temperature, are suitable, the familiar toadstool appears, it being only the reproductive stage of the fungus. Its rate of growth is often phenomenal, as is also the force it may exert in its growth. Seemingly delicate bodies not only break open the hard-packed surface of the ground, but also may push aside pebbles of considerable weight. Not infrequently whole rings of toadstools appear in a field, springing to maturity in a single night—these are the familiar fairy rings, resulting from the growing outward from a common source of the unseen mycelium.

When the young fruit-body first comes up it is completely enclosed in a membranous skin known as the velum. As enlargement continues this skin is broken. Often traces of the velum remain in the form of flakes on the upper surface of the cap, and as a ring or annulus around the stalk. Attempts have been made to find in these characteristics a means for separating the edible from the poisonous species. However, no reliable distinction is found here. Actually, unless one is absolutely sure of the identity of a given species, the only safe rule is complete abstinence. Classification is based on the color of the spore-masses, and also the determination of the way in which the gills are attached to the stalk, the color changes shown as a cut or broken surface dries, and various other means.

A few species of Agarics, notably *Agaricus (Psalliota) campestris*, are extensively cultivated, and justly esteemed as food. Other species are violently poisonous, particularly species of the genus *Amanita*. Yet certain Siberian people use *Amanita muscaria*, a poisonous species, to produce a form of intoxication. Another common mushroom, the Inkycap, a species of *Coprinus*, has been used as a writing fluid, the substance of the toadstool breaking down into a fluid mass containing vast numbers of black spores. See **Basidiomycetes.**

Mushrooms are cultivated extensively in certain regions, usually in caves or specially constructed buildings. "Spawn" is prepared by inoculating grain, commonly rye, with spores or mycelium of the desired variety, and allowing the fungus to grow for a time. The spawn is spread as "seed" upon beds of compost where the fungus develops the characteristic mycelium described above. The compost is prepared from horse manure, including the bedding straw, from ground corn cobs, hay, or other organic materials. After the mycelium is established, the bed is covered by a thin layer of carefully selected soil. Usually thousands of mushrooms break through the soil at one time, and are harvested at the proper stage of growth. The entire growth takes place in the dark under controlled temperature and humidity.

Additional Reading

Clay, C.: "The Finest Fungus (Mushroom) Among Us—and How to Find Them," *Amer. Forests*, 34 (March/April 1991).

Krieger, G.: *The Mushroom Handbook*, Dover Publications, New York, 1967.

Lincoff, G., and D. H. Mitchel: *Toxic and Hallucinogenic Mushroom Poisoning*, Van Nostrand Reinhold, New York, 1977.

Rinaldi, A., and V. Tyndalo: *The Complete Book of Mushrooms*, Crown Publishers, New York, 1974.

Staff: "Stalking Wild Mushrooms," *Nat'l. Geographic*, 138 (May 1991).

AGATE. Agate is a variety of chalcedony, whose variegated colors are distributed in regular bands or zones, in clouds or in dendritic forms, as in moss agate.

The banding is often very delicate with parallel lines of different colors, sometimes straight, sometimes undulating or concentric. The parallel bands represent the edges of successive layers of deposition from solution in cavities in rocks which generally conform to the shape of the enclosing cavity.

As agate is an impure variety of quartz it has the same physical properties as that mineral. It is named from the river Achates in Sicily where it has been known from the time of Theophrastus.

Agate is found in many localities; India, Brazil, Uruguay, and Germany are notable for fine specimens.

Onyx is a variety of agate in which the parallel bands are perfectly straight and can be used for the cutting of cameos. Sardonyx has layers of dark reddish-brown carnelian alternating with light and dark colored layers of onyx.

See also **Chalcedony;** and **Quartz.**

AGAVE (*Amaryllidaceae*). A large genus of plants, particularly abundant in Mexico, in which the thick rigid leaves form a basal rosette from the center of which rises the tall flower stalk. Because of the time required to store sufficient food reserves for flowering, certain species, notably *Agave americana*, are called century plants from the belief that they flower but once a century; actually flowering may occur in from 5 to 50 or more years. Once started, the flower stalk develops very rapidly, requiring immense quantities of sap. In Mexico, the flower stalk is cut off early in its formation and the stump scooped out to form a cup into which quantities of sweet sap exude. This sap is collected and fermented to form pulque, a strong drink with an unpleasant odor. Distilled pulque gives a more potent drink, mescal. From the leaves of several species, particularly *Agave sisalana* and *A. fourcroydes*, are obtained fibers. These fibers occur as sclerenchyma sheathes surrounding the vascular bundles in the leaves. To obtain the fibers, the leaves are cut off and the spiny tip and margin removed. Machines then heat and scrape the leaves and wash them until the clean fibers are obtained. These are then dried either in the sun or by artificial heat, and are ready for export under the name of sisal or henequen, according to the species from which they were obtained. Many species of *Agave* are cultivated for their ornamental value. See **Century Plant.**

AGENT ORANGE. Common name for a 50–50% mixture of the herbicides 2,4,5-T and 2,4-D, once widely used by the military as a defoliant. The mixture contains dioxin as a contaminant. See also **Dioxin;** and **Herbicide.**

AGGLOMERATE. A term proposed by Sir Charles Lyell in 1831 for coarsely graded volcanic ejectamenta similar in appearance to ordinary conglomerates or breccias. An extremely thick and widespread accumulation of so-called agglomerates occurs on the borders of the Yellowstone Park. These deposits, however, include numerous beds of water-laid pebbles, gravels and sands, the latter containing fossil plants of early tertiary age.

Because of the various interpretations of the term, it should be defined in context.

AGGLOMERATION. This term connotes a gathering together of smaller pieces or particles into larger size units. This is a very important operation in the process industries and takes a number of forms. Specific advantages of agglomeration include increasing the bulk density of a material, reducing storage-space needs, improving the handling qualities of bulk materials, improving heat-transfer properties, improving control over solubility, reducing material loss and lessening of pollution, particularly of dust, converting waste materials into a more useful form, and reducing labor costs because of resulting improved handling efficiency.

The principal means used for agglomerating materials include (1) compaction, (2) extrusion, (3) agitation, and (4) fusion.

Tableting is an excellent example of compaction. In this operation, loose material, such as a powder, is compressed between two opposing surfaces, or compacted in a die or cavity. Some tableting machines use the action of two opposing plungers which operate within a cavity. Resulting tablets may range from $\frac{1}{8}$ to 4 inches (3 millimeters to 10 centimeters) in diameter. Uniformity and dimensional precision are outstanding. Numerous pharmaceutical products are formed in this manner, as well as some metallic powders and industrial catalysts.

Pellet mills exemplify the use of extrusion. In some designs the charge material is forced out of cylindrical or other shaped holes located on the periphery of a cylinder within which rollers and spreaders force the bulk materials through the openings. A knife cuts the extruded pellets to length as they are forced through the dies.

The *rolling drum* is the simplest form of aggregation using agitation. Aggregates are formed by the collision and adherence of the bulk particles in the presence of a liquid binder or wetting agent to produce what essentially is a "snowball" effect. As the operations continues, the spheroids become larger. The strength and hardness of the enlarged particles is determined by the binder and wetting agent used. The operation is followed by screening, with recycling of the fines.

The *sintering process* utilizes fusion as a means of size-enlargement. This process, used mainly for ores and minerals and some powdered metals, employs heated air which is passed through a loose bed of finely ground material. The particles partially fuse together without the assistance of a binder. Sintering frequently is accompanied by the volatilization of impurities and the removal of undesired moisture.

The *spray-type* agglomerator utilizes several principles. Loosely bound clusters or aggregates are formed by the collision and coherence of the fine particles and a liquid binder in a turbulent stream. The mixing vessel consists of a vertical tank, around the lower periphery of which are mounted spray nozzles for introduction of the liquid. A suction fan draws air through the bottom of the tank and creates an updraft within the mixing vessel. Materials spiral downward through the mixing chamber, where they meet the updraft and are held in suspension near the portion of the vessel where the liquids are injected. The liquids are introduced in a fine mist. Individual droplets gather the solid particles until the resulting agglomerate overcomes the force of the updraft and falls to the bottom of the vessel as finished product.

AGGLUTINATION. 1. The gathering of particles. 2. The clumping together of bacteria or cells, resulting often from their reaction with the corresponding immune or modified serum.

AGGLUTININ. One of a class of substances found in blood to which certain foreign substances or organisms have been added or admixed. As the name indicates, agglutinins have the characteristic property of causing agglutination, especially of the foreign substances or organisms responsible for their formation.

AGGRADATION. In geology, the building up of the surface of the earth by deposition, as of sediment by a river in its valley. More specifically, the upbuilding caused by a stream so as to establish and maintain a uniform grade or slope. The term also is used sometimes as a synonym for accretion, as in the case of development of a beach. See also **Accretion; Alluvial Fan.**

AGGREGATE. The solid conglomerate of inert particles which are cemented together to form concrete are called aggregate. A well-graded mixture of fine and coarse aggregates is used to obtain a workable, dense mix. The aggregate may be classed as fine or coarse depending upon the size of the individual particles. The specifications for the concrete on any project will give the limiting sizes which will distinguish between the two classifications. Fine aggregate generally consists of sand or stone screenings while crushed stone, gravel, slag or cinders are used for the coarse aggregate. The aggregates should be strong, clean, durable, chemically inert, free of organic matter, and reasonably free from flat and elongated particles since the strength of the concrete is dependent upon the quality of the aggregates as well as the matrix of cementing material. See also **Concrete.**

AGGRESSIN. A product of bacterial metabolism which impairs the defensive mechanisms of the blood of the host.

AGING. See **Gerontology and Geriatrics.**

AGRANULOCYTOSIS. This is a potentially serious syndrome in which the white cells may be greatly decreased or almost absent from circulation. Because the granulocytes are important in protecting the body against infection, an individual deprived of these defensive forces for long may have an overwhelming invasion of the bloodstream and organs with dangerous disease-producing organisms. See also **Blood.** Important symptoms include general weakness, prostration, headache, shaking chills, and progressive ulcerative throat lesions. Diagnosis must be confirmed by studies of the bone marrow.

The syndrome of agranulocytosis may result from an overwhelming infection, releasing toxins specifically destructive to the bone marrow and lymphatic systems. It may develop secondary to allergic sensitization to drugs with antigenic properties, such as aminopyrine, cimetidine, potassium-sparing diuretics, procainamide, propanolol, and sulfapyridine. Industrial toxins, particularly benzol, may damage the marrow similarly.

AGREEMENT (Coefficient of). This coefficient relates to the situation where m observers each provide paired comparisons for n objects. A coefficient of agreement between the verdicts of the m observers is given by

$$u = \frac{8\Sigma}{m(m-1)(n-1)n} - 1$$

where Σ is the sum of the number of agreements between pairs of judges.

The coefficient may vary from $-1/(m-1)$, if m is even, or $-1/m$, if m is odd, up to $+1$ if there is complete agreement.

AGROSTOLOGY. The science of grasses, including classification, management, and utilization. See **Grasses.**

AGULHAS CURRENT (also called Agulhas Stream). A generally southwestward-flowing ocean current of the Indian Ocean; one of the swiftest of ocean currents.

Throughout the year, part of the south equatorial current turns south along the east coast of Africa and feeds the strong Agulhas current. To the south of latitude 30°S, the Agulhas current is a well-defined and narrow current that extends less than 100 kilometers (62 miles) from the coast. To the south of South Africa, the greatest volume of its waters bends sharply to the south and then toward the east, thus returning to the Indian Ocean by joining the flow from South Africa toward Australia across the southern part of that ocean. However, a small portion of the Agulhas current water appears to round the Cape of Good Hope from the Indian Ocean and continue into the Atlantic Ocean.

AIDS. See **Immune System and Immunology.**

AIChE. The American Institute of Chemical Engineers was founded in Philadelphia, Pennsylvania, in 1908 to serve which, at that time, was an emerging new engineering discipline, *chemical engineering*. The general aim of the Institute is to promote excellence in the development and practice of chemical engineering through semiannual district meeting and an annual national meeting for the presentation and discussion of technical papers and the exhibition of equipment and materials used in chemical engineering projects. The Institute publishes several periodicals, including the *AIChE Journal, International Chemical Engineering,* and *Chemical Engineering Progress*. Technical divisions of the AIChE include Computer and Systems Technology, Engineering and Construction Contracting, Environmental Technology, Food, Pharmaceutical and Bioengineering, Forest Products, Fuels and Petrochemicals, Heat Transfer and Energy Conversion, Management, Materials Engineering and Sciences, Nuclear Engineering, Safety and Health, and Separations Technology. The Institute sponsors research projects in cooperation with corporate, governmental, and institutional sources, including the Center for Chemical Process Safety (CCPS), the Center for Waste Reduction Technologies (CWRT), the Design Institute for Emergency Relief Systems (DIERS), the Design Institute for Physical Property Data (DIPRR), the Process Data Exchange Institute (PDXI), and the Research Institute for Food Engineering. Headquarters of the AIChE is in New York City.

AILERON. The aileron is one of the three aerodynamic surfaces of an airplane which are variable in position at the will of the pilot, the purpose of which is to provide the required degree of maneuverability and control of the aircraft about its longitudinal axis. The aileron is that surface which causes the necessary forces to be produced to induce rotation of the aircraft.

This motion is known as *roll*, which is employed to correct other rolls produced unintentionally, as by gusts—when not used to accomplish such maneuvers as banks or sideslips.

The conventional type of aileron is a flap inserted in the trailing edge of the wing, usually at the tip. This flap is rotatable around its forward axis, upward and downward, and in effect changes the camber of the airfoil. The result is a change of pressure on that portion of the wing over which the aileron extends. The change in pressure is greater on the side where the aileron is rotated downward and less on the side where it is rotated upward. The difference in pressure causes the airplane to roll. When the airplane has reached the required angle of bank, the ailerons are returned to neutral. The ailerons are then used in the opposite direction to bring the airplane to its level or normal position, when again the ailerons are returned to neutral. The principal defect of the flap-type aileron is that it becomes relatively ineffective when, for safety's sake, it is needed to be most effective, i.e., when the airplane approaches the stall, or is stalled. (This stall situation refers to the aerodynamic stall of the wing, not to the stall of the engine.) With adequate control of roll during a stall, many serious accidents involving tailspin can be avoided. This defect can be minimized by careful design, and avoided by proper handling of the aircraft. The flap-type ailerons produce some adverse yawing moments, in addition to the desirable rolling moments. This yawing moment is counteracted by use of the rudders.

The aerodynamic efficiency, simplicity, and reliability of the flap-type aileron are better than for other types, such as wing-tip ailerons and spoiler ailerons. See also **Aerodynamics and Aerostatics.**

AIMLESS DRAINAGE. A type of drainage or stream pattern that occurs in low swampy lands; particularly characteristic of glaciated regions of low relief. Essentially, drainage without a well-developed system.

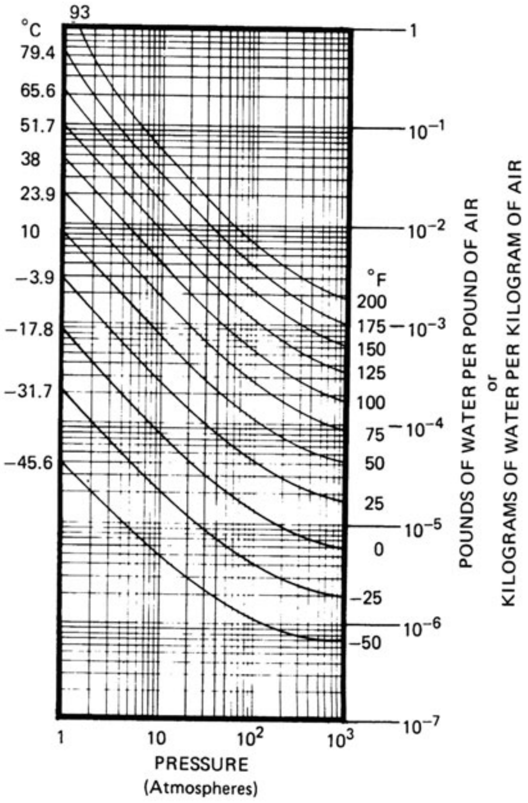
AIR. In addition to being the principal substance of the earth's atmosphere, air is a major industrial medium and chemical raw material. The average composition of dry air at sea level, disregarding unusual concentrations of certain pollutants, is given in Table 1. The amount of water vapor in the air varies seasonally and geographically and is a factor of large importance where air in stoichiometric quantities is required for reaction processes, or where water vapor must be removed in air-conditioning and compressed-air systems. The water content of air for varying conditions of temperature and pressure is shown in Table 2. The water content of saturated air at various temperatures is shown in the accompanying figure. See also **Oxygen; Nitrogen;** and **Pollution (Air).**

TABLE 1. COMPOSITION OF AIR

Constituent	Percent by Weight	Percent by Volume
Oxygen (O_2)	23.15	20.95
Ozone (O_3)	1.7×10^{-6}	0.00005
Nitrogen (N_2)	75.54	78.08
Carbon dioxide (CO_2)	0.05	0.03
Argon (Ar)	1.26	0.93
Neon (Ne)	0.0012	0.0018
Krypton (Kr)	0.0003	0.0001
Helium (He)	0.00007	0.0005
Xenon (Xe)	5.6×10^{-5}	0.000008
Hydrogen (H_2)	0.000004	0.00005
Methane (CH_4)	trace	trace
Nitrous oxide (N_2O)	trace	trace

TABLE 2. WATER CONTENT OF SATURATED AIR

Temperature		Water Content (Pounds in 1 Pound of Air, or Kilograms in 1 Kilogram of Air)
(°F)	(°C)	
40	4.44	0.00520
45	7.22	0.00632
50	10	0.00765
55	12.8	0.00920
60	15.6	0.01105
65	18.3	0.01322
70	21.1	0.01578
75	23.9	0.01877
80	26.7	0.02226
85	29.4	0.02634
90	32.2	0.03108
95	35.0	0.03662
100	37.8	0.04305
105	40.6	0.05052



Water content of saturated air at various temperatures and pressures.

AIR AND OTHER GAS COMPRESSION. The compression of air by mechanical means, and the raising of it to some desired pressure above that of the atmosphere, is effected, usually, by an approximate adiabatic change of state.

If the ideal compression were possible, it would be represented by the following equation showing the relation between pressure and volume:

$$PV^{1.4} = \text{a constant}$$

A compression of this nature may heat the air to temperatures which would interfere with reliable action of an air compressor and introduce lubrication difficulties, were there no provision for cooling the cylinder walls. Therefore, in compressors we find the cylinders to be externally finned or water-jacketed so that sufficient cooling is secured to keep the temperatures from becoming excessive. The extraction of heat from the cycle in this way modifies the conditions of compression from the ideal to some change more nearly represented by

$$PV^n = C$$

in which n usually lies between 1.35 and 1.4. The ratio of the temperature before and after compression is expressed by the following equation, the temperatures being degrees Fahrenheit absolute.

$$\frac{T_2}{T_1} = \left[\frac{V_1}{V_2}\right]^{n-1}$$

In compression to high pressures, the temperature rise may be too great to permit the compression to be carried to completion in one cylinder, even though it is cooled as mentioned above. In high-pressure compressors, the compression is carried out in stages, with a partial increase of the pressure in each stage, and cooling of the air between the stages. Two- and three-stage compression is very common where pressures of 300–1000 lbs. per sq. in. (20–68 atmospheres) are needed.

The volume of clearance air should be made as small as possible in order to improve the volumetric efficiency of the compressor, since the clearance air must expand to the suction pressure before the cylinder can begin to be charged.

The mechanical construction of air compressors varies with the amount of compression required and the character of service.

Air Compressors. Mechanical air compressors can be classified into two major groups (1) *reciprocating compressors* of the piston-and-cylinder type; and (2) *rotating compressors*, which may be further divided into a number of kinds. Since the reciprocating compressors were the earliest type to be developed, at any rate for higher pressors, they are discussed first in this entry.

Reciprocating Compressors. A common type of air compressor is the piston and cylinder compressor in which a reciprocating piston positively displaces the air from a cylinder during its discharge stroke. Compressors for charging tanks of air used to inflate pneumatic tires at the numerous automotive service stations are of the reciprocating type. Being of small capacity they are generally single acting and air cooled (by exterior fins) since those features are common in small compressors. Larger compressors, unless for extremely high pressure, are usually double acting and frequently cooled by water jackets. One or two cylinder arrangements are conventional, with a tendency to secure large capacity by increased bore and stroke rather than by multiple cylinders. Pistons are reciprocated by a crankshaft and connecting rod mechanism commonly deriving motion from the driving source by belt. Valves are springloaded to open upon slight differential pressures.

The compressor cycle is shown in Fig. 1. The discharge stroke which begins at A builds up the pressure to B where it exceeds the receiver pressure sufficiently to open a discharge valve. Discharge then takes place at a constant control pressure from B to C. The volume C is the clearance volume of the compressor. The air in the clearance volume must expand to D during the suction stroke before the inlet valve will open. Thus the volume of air drawn in per stroke is only that from D to A. Obviously the compressor should have as small a clearance as possible in order to obtain good volumetric efficiency, especially at high discharge pressures. Small compressors may be operated with high compression ratios (8–12) if desired because cooling is more effective in small cylinders and mechanical strength is readily provided. Large volumes compressed to ratios exceeding 4 will need a multi-stage compressor to permit cooling between stages and to lessen the structural loads on the large first stage cylinders.

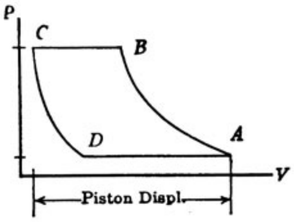


Fig. 1. Reciprocating compressor cycle.

Rotating compressors may be classified into four types ranging from the (1) centrifugal and (2) axial compressors that are often used for high pressure service to (3) blowers and (4) fans used for low pressure—large volume operation. Since, however, the term compressor is so often restricted to higher pressure service, blowers and fans are treated in separate entries in this book, leaving this one to deal with centrifugal and axial flow compressors.

Centrifugal Compressors. A rotating impeller mounted in a casing and revolved at high speed will cause a fluid which is continuously admitted near the center of rotation to experience an outward flow and a pressure rise due to centrifugal action.

Assume that an impeller with radial blades of depth $r_2 - r_1$ is revolving at a speed of ω radians per minute. This is illustrated in Fig. 2. Consider that a compressible fluid (a gas) is admitted at the center and flows into the impeller radially. Relative to the impeller blades it has an outward radial flow, finally emerging with some absolute velocity v_2 which is partially diffused into pressure. In addition a pressure gradient must exist to balance the sum of all the incremental $mr\omega^2$ inertia forces arising from the inward acceleration $r\omega^2$ given each particle of the fluid. The interrelation of r, ω, P, can readily be developed by considering the power required as (1) that necessary for the thermodynamics of compression, and (2) that which would account for the action of the impeller in effecting certain momentum changes on the fluid.

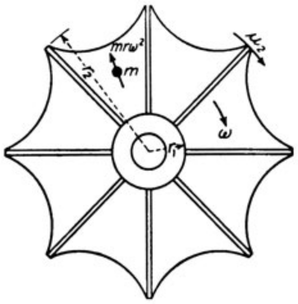


Fig. 2. Impeller for centrifugal compressor.

Without detailing the algebraic procedure, it may be stated that the ability of an impeller to raise the fluid pressure is expressed by the ratio of discharge to inlet pressure, i.e., P_2/P_1

$$\frac{P_2}{P_1} = \left[1 + \frac{\eta z u_2^2}{gRT_1}\right]^{1/z}$$

$g = 32.2$
$R = 53.4$ for air
$z =$ a gas coefficient, about 0.286 for air
$T_1 =$ energy coefficient. This would be 1.0 if the flow through the impeller were non-turbulent and frictionless. Typically, $\eta = .75$ to $.85$.
$u_2 =$ impeller rim velocity, ft per sec.

For pressure ratios higher than can be obtained from the action defined above, several impellers may be mounted on the same shaft and enclosed in a compound casing with passages arranged to lead the output from one impeller to the "eye" of the next. This multistaging principle is used to produce pressures above the capability of a single impeller compressor. Multi-stage compression may have cooling between the stages so the overall compression may be more isothermal than adiabatic. If compression were isothermal

$$\frac{P_2'}{P_1} = e^{\eta u2/gRT}$$

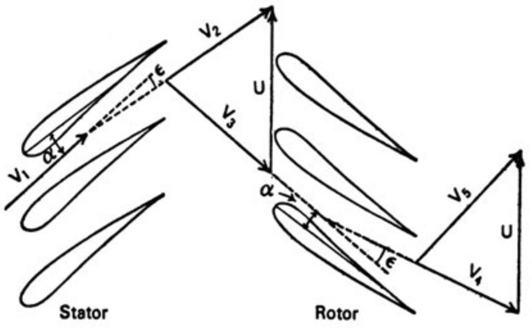


Fig. 3. Velocity relations in axial-flow compressing blading.

Axial-flow Compressors. Figure 3 illustrates the compressing action in an axial-flow multistage compressor. Air flows over a set of stationary airfoils (arranged circumferentially as fixed blading) with an angle of attack α. As the airfoil aspect ratio is small, the downwash created is considerable. Downwash turns the air stream through angle ϵ, which results in a flow channel of increasing cross section. The diffusion thus effected slows down the air velocity to V_2 and increases pressure. The moving airfoils (rotor blades) receive this air stream at V_3 as a result of the vectorial combination of air stream velocity V_2 and blade speed U. Relative to the moving airfoils the downwash again causes a diffusion of velocity and another increase of pressure. The relative velocity V_4 leaving the moving blades produces an absolute final velocity of V_5 because of blade motion. Suitable combinations of U and α enable V_5 to duplicate V_1 and permit a similar following stage. The blade height in the following stage is decreased because of the smaller specific volume of the compressed air. Increasing α will increase the pressure rise without much decrease in volume. The axialflow compressor should be designed for conservative α's for two reasons. First, an operation near the stalling angle would be hazardous as small variations of α might occur which could burble the airfoils and cause an unstable, rough, or even hazardous condition to exist. Secondly, axial-flow compressors may be employed under conditions where utmost efficiency is imperative (as in gas turbine power units) and α should create the optimum favorable balance between good downwash and minimum turbulent airfoil wake. Axial-flow compressors have been built with energy efficiences as high as 90%. They may be operated effectively at high speeds, i.e., 5000–50,000 rpm.

The compression of air and other gases may also be secured by the employment of steam jets if an admixture of vapor in the compressed gas is not undesirable. High-pressure steam is blown through nozzles which create a high-velocity jet. The gas to be compressed is led into the regions about the nozzle discharge where it is entrained in the steam jet. The mixture then travels into a diffuser for compression and attendant velocity reduction. Although the compression thus achieved is of limited magnitude, staging the compression in a series of nozzles, with intermediate coolers for partial condensation of vapor, allows moderate compression ratios to be achieved.

See also **Gas and Expansion Turbines.**

Screw Compressors. Although invented in the 1930s, the oil-injected screw compressor was not considered seriously by the oil and chemical industries until the 1960s. Today, screw compressors are used widely in refrigeration and natural gas compression. These compressors are oil-injected helical screw machines that are positive-displacement devices. See Fig. 4. In this compressor, a volume of gas is trapped in a defined space. The volume is reduced with a corresponding pressure increase, as in a reciprocating compressor. As pointed out by B. C. Price (Arco Oil & Gas Co.), "The difference is that this compression action occurs without reciprocating motion and the resulting pulsations. Gas is transported axially through the machine in a continuous flow without pulsations. Compression is accomplished without the need for suction and discharge valves and thus has fewer moving parts."

The screw compressor consists of two rotors enclosed in a pressure-containing casing. The rotors differ in shape and are identified as the "male" and "female" unit. The male rotor has four lobes with an asymmetric profile formed helically along the shaft. These lobes mesh with the six flutes of the female rotor. As the units rotate, gas is drawn through the suction nozzle into the interlobe space. As rotation continues, the rotors mesh to trap the gas and isolate it from the suction. The volume of gas constantly is reduced along its path, with a resulting increase in pressure. While the gas is being moved along axially, the next charge of gas is drawn into the machine and the process is repeated.

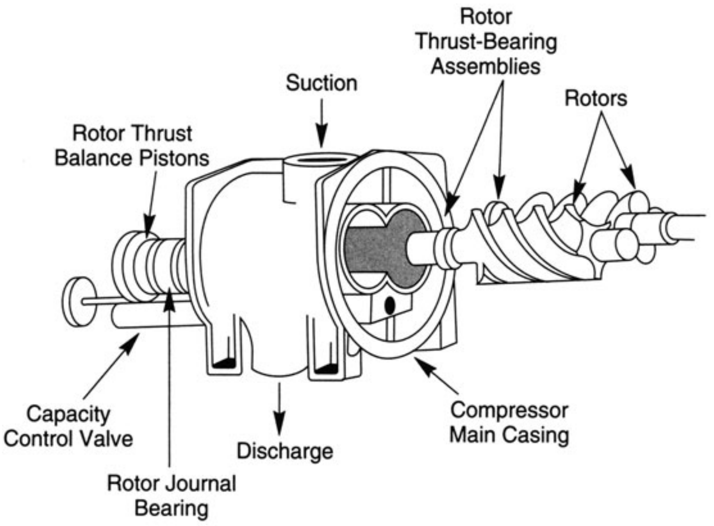


Fig. 4. Schematic of screw-type gas compressor. (*After Price.*)

Further details on the machine can be found in an article by B. C. Price (*Chemical Eng. progress*, 50, February 1991).

AIRCRAFT ICING. Aircraft flying in airspace that is also occupied by liquid water drops and droplets at temperatures below 0°C experience ice accumulation on the leading edges and surfaces of their structures. Water can remain in liquid form to temperatures near −40°C although most clouds tend to be composed of ice crystals in the temperature range of −20 to −30°C or lower. The airspace in which aircraft icing is most commonly encountered lies between the earth's surface and 20,000 feet (6000 meters).

There are several factors involved in aircraft icing. The amount of supercooled water in the space swept out by the moving plane determines in major part the intensity of icing, i.e., the rate of accumulation. The speed of the plane is another factor because at high speeds the plane sweeps out more space in less time. Drop and droplet size has a modifying effect on accumulation. Very small droplets tend to follow the airflow around the plane's surfaces and do not impact to freeze. Large drops like rain tend to abandon the airstream, impact on the plane and freeze. A cloud with large numbers of very small droplets therefore may be much less of a hazard than airspace containing many fewer drops of falling rain.

The temperature of the airspace determines in large part the type of icing. When drops and droplets freeze instantly the water material does not have time to assume a crystalline structure and the ice formation tends to be opaque, often granular with occluded air, usually less dense than clear ice, and frequently an irregular surface. This is *rime ice.* On the other hand, when liquid droplets and drops do have time to crystallize, the ice formation tends to be more or less clear, usually near the denisty of ordinary ice and tends to assume the shape of the external aircraft surfaces over which it spreads as it freezes. This is *clear ice.* Temperatures just less than freezing are conducive to clear icing and temperatures well below 0°C are associated with rime icing. See Figs. 1 and 2. Aircraft icing is usually a mix of the two types with the proportion of the two related to the temperature.

Clouds most prone to contain an icing hazard are those of cumulo form including billowing stratocumulus. Clouds least prone to contain icing or to contain icing of lesser intensity are stratus. Rain falling from a warm layer of air into a cold lower layer whose temperature is less than 0°C is the most intense of all icing conditions. Meteorological arrangements that produce freezing rain or supercooled rain are predomi-

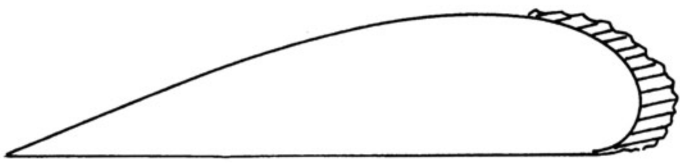


Fig. 1. Rime ice accumulation on the leading edge of a wing tends to be irregular, rough, and a spoiler of the air flow, but is not as heavy as clear ice.

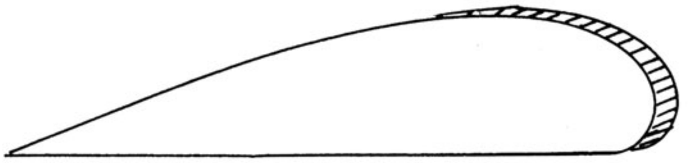


Fig. 2. Clear ice accumulation on the leading edge of a wing tends to be smooth and symmetrical, but heavy.

nantly along the cold side of warm fronts and secondarily behind cold fronts. Aircraft icing remains a very serious problem.

For references, see entries on **Climate;** and **Meteorology.**

Peter E. Kraght, Certified Consulting Meteorologist,
Mabank, Texas.

AIR CONDITIONING. See **Heat Pump; Solar Energy.**

AIR GAP. This term is commonly used in connection with various magnetic circuits and denotes a gap left in the magnetic material. In the construction of various chokes and transformers used in communications circuits a short gap is usually left in the core material to prevent saturation of the magnetic material by the d.c. which often flows in the one or more windings wound on the core material. See **Electromagnet.**

In rotating electrical machinery the rotating part of the magnetic circuit must, of course, be separated by a gap. In these machines this gap is kept as small as is consistent with adequate mechanical clearance. In most instances the gap introduces no desirable electrical or magnetic characteristics and necessitates the application of additional electrical energy to overcome its reluctance. The term air gap is sometimes used to denote any spark gap, comprising two electrodes separated by a fluid, commonly air.

AIR LIFT PUMP. An air lift is a water pumping method whereby water may be raised from a well through the medium of compressed air. The drop pipe in the well is supplied at the bottom with compressed air from a small air pipe, and the effect of mixture of air and water at the bottom of the drop pipe is to bring water to the surface. This is accomplished either by the water acting as pistons, trapping intermediate layers of air, the expansion of which drives the water pistons to the surface; or by the mingling of air and water, forming a mixture that is sufficiently lighter than the undisturbed water in the well so that the mixture rises above the surrounding water. In order for this rise to reach the surface the discharge pipe must be submerged in the water of the well an amount varying from 100–300% of the actual lift. The pumping efficiency of the system is very low, but it is very suitable for handling gritty or corrosive waters.

AIR LOCK. An air lock is an air-tight compartment in which the air pressure may be regulated to any desired intensity. When workers are required to work in regions where the air pressure is above that of the atmosphere, an air lock must be provided to permit passage of the workers from the open atmosphere to the pressure region. Thus, in the case of caissons, where workers must labor under a high enough pressure to equalize the hydrostatic pressure existing at the bottom of the caisson, or in tunnels where flooding is avoided by forcing compressed air into it at sufficiently high pressures to hold the water back, the air lock is a feature essential to the maintenance of pressure during the admission of workers. Separate air locks are generally provided for materials and spoil removal. In construction work, the air lock is a chamber of sufficient size to hold the number of persons that must be accommodated in it at one time. It is provided with well-braced doors having sealing-type edges and tightening locks. The chamber is equipped with valves for admitting and releasing air and with safety devices to prevent excessive pressures endangering the lives of the occupants. The air lock must have two air-tight doors, one leading to the atmosphere, the other leading to the pressure region. These doors open inward so that the pressure in the air lock tends to tighten them against the frame. To enter a pressure region, a caisson for example, the workers enter the air lock, after which the door leading to the atmosphere is tightly fastened. Compressed air is then slowly admitted until the pressure in the air lock equals that in the caisson, after which the connecting door may be opened without trouble or loss of air from the working chamber. After the workers enter the caisson, the air lock door is tightly closed, after which the air lock may be opened from the outside without affecting conditions within the caisson.

AIR MASS. A very large bulk of air, having properties (temperature, humidity, thermal structure) that vary only slightly, or vary linearly from point to point within the parcel. Air masses range from about 500–5000 miles (805–8045 kilometers) in lateral dimensions. They develop over large, relatively homogeneous, geographical areas where air is stagnant for a sufficient period to acquire the characteristics of that region. These regions are either continental or maritime and are known as air-mass source regions. After an air mass begins to move from its source region, it acquires modifying features characteristic of the surface over which it travels. Modification continues until the air mass loses its identity in the general atmospheric circulation.

A number of systems have been proposed for the classification of air masses, but the Bergeron classification has been the most widely accepted. In this system, air masses are designated first according to the thermal properties of their source regions: tropical (T); polar (P); and less frequently, arctic or antarctic (A). For characterizing the moisture distribution, air masses are distinguished as to continental (c) and maritime (m) source regions. Further classification according to whether the air is cold (k) or warm (w) relative to the surface over which it is moving indicates the low-level stability conditions of the air, the type of modification from below, and is also related to the weather occurring within the air mass. This outline of classification yields the following identifiers for air masses: cTk, cTw, mTk, mTw, cPk, cPw, mPk, mPw, cAk, cAw, mAk, mAw.

Other classification systems introduce further distinction between stable (s) and unstable (u) conditions in upper levels. Some include equatorial (E), monsoon (M), or superior (S) air. Others omit the arctic (A) type and classify all air masses on the basis of polar and tropical air, separated by the polar front.

The major air classification types may be defined as follows:

Tropical air is developed over low latitudes. Maritime tropical air (mT), the principal type, is produced over the tropical and subtropical seas. It is very warm and humid, and is frequently carried poleward on the western flanks of the subtropical highs. Continental tropical air (cT) is produced over subtropical arid regions, and is hot and very dry.

Polar air is developed over high latitudes, especially within the subpolar highs. Continental polar air (cP) has low surface temperature, low moisture content, and, especially in its source regions, great stability in the lower layers. It is shallow in comparison with arctic air. Maritime polar air (mP) initially possesses similar properties to those of continental polar air; but in passing over warmer water, it becomes unstable with a higher moisture content.

Arctic air is developed mostly in winter over arctic surfaces of ice and snow. It is cold aloft and extends to great heights, but the surface temperatures are often higher than those of polar air. For two or three months in summer, arctic air masses are shallow and rapidly lose their characteristics as they move southward.

Continental air is developed over a large land area, and has the basic continental characteristic of relatively low moisture content.

Maritime air is developed over an extensive water surface, and has the basic maritime quality of high moisture content in at least its lower levels.

Equatorial air, according to some classifications, is the air of the doldrums, or the equatorial trough, to be distinguished somewhat vaguely from the tropical air of the trade-wind zones. Tropical air "becomes" equatorial air when the former enters the equatorial zone and stagnates. There is no significant distinction between the physical properties of these two types of air in the lower troposphere.

Superior air is an exceptionally dry mass of air usually found aloft but occasionally reaching the earth's surface during extreme subsidence processes (that is, the descending motion of air in the atmosphere, usually with the implication that the condition extends over a rather broad area). It is often found above tropical maritime air.

For references see entries on **Climate;** and **Meteorology.**

Peter E. Kraght, Certified Consulting Meteorologist,
Mabank, Texas.

AIRPLANE. The design and construction of a practical, top-performing aircraft is one of the most demanding of engineering feats. Engineers must wrestle constantly to achieve optimal solutions to inherently conflicting factors, — such as aircraft speed versus fuel economy, range versus speed, controllability versus complexity, weight versus strength of materials, passenger comfort and cargo handling efficiency versus initial costs, and the list goes on and on. In addition, all such choice decisions must be made without compromising safety. In some engineering fields, such as bridge or dam building, reasonably optimal design decisions have been made with over a century of experience. Although aviation spans nearly a century from the Wright brothers and Kitty Hawk (1903), design experience simply has not been one of linear extrapolation of initial expertise, but rather airplane design trends have been interrupted periodically by radical departures from the past, as exemplified by changing concepts of motive power (jet versus reciprocating engines) and vast improvements in fuels, or by improvements in materials of construction (new metal alloys and composites), or by new findings pertaining to the very fundamentals of aerodynamics, or by the availability of grossly improved instrumentation and control that has fallen out of solid-state physics and electronics. Again, the list could go on and on. Further, the basic function or need for aviation also has changed over the years, notably including the autogyro and later the helicopter, as well as the transformation of an aircraft to an indispensable tool of the military and the employment of aircraft for scientific reconnaissance. Thus, a search of the history of aircraft design exhibits comparatively short periods of design stability that are punctuated every few years by fundamental changes.

For example, the introduction of large, jet-powered aircraft, such as the DC-10 with its first flight in August 1970, ushered in an era of some stability in passenger- and cargo-carrying aircraft, punctuated by continuing improvements to a fundamental design. This was reflected by the frequent introduction of new model numbers by the principal manufacturers, including McDonnell Douglas and Boeing, as well as a few European aircraft manufacturers. See Figures 1 through 4. The next era in this field commenced in the late 1980s and early 1990s with the introduction by McDonnell Douglas of the MD-11 tri-jet, the MD-11 COMBI, and the Boeing 777 (delivery scheduled for 1995).

If past history is any guide, these are the kinds of airplanes that, again with subsequent improved models, will carry the technology well past the year 2000 (most likely to 2020), at which time another new era will commence. With federal support funding very tight as of 1994, it appears unlikely that, with the exception of continued research, the hyper-

Fig. 1. The DC-10 ushered in the aircraft design era of the 1970s. For nearly 25 years, this has been a virtual workhorse of passenger aviation. The first commercial flight took place in August 1970. The DC-10, with many still in use, is powered by three engines ranging in thrust from 40,000 to 53,000 pounds (178 to 236 kN).

The length is approximately 182 feet (55.4 meters), wing span is over 165 feet (50 meters) on some models; the height is just over 58 feet (17.6 meters). Maximum gross weights, depending upon model, range from 440,000 to 572,000 pounds (199,584 to 259,460 kilograms). Maximum number of passengers carried is up to 380 persons, depending on interior arrangement. Depending upon model, the range for a typical mixed class is from approximately 4300 to 6180 miles (6920 to 9940 kilometers). Cruise speed is 510 knots or 587 miles (945 kilometers) per hour.

As of late 1986, over 440 DC-10s and KC-10s had been produced. The KC-10 is an advanced tanker/cargo aircraft, based upon the DC-10 Series 30CF (convertible freighter). The fleet of DC-10s has flown more than 5.3 billion miles (8.5 billion km) worldwide. Fifty airlines operate DC-10s, serving more than 200 cities. Each week, DC-10s make more than 7000 flights, carrying more than 1.1 million passengers. As of September 1986, DC-10s had transported over 605 million passengers. Revenue flying time totaled more than 11.4 mil hours, passenger miles surpassed 842 billion (1355 billion passenger kilometers), and landings, 3.86 million. Statistics complied as of June 1, 1986 place the average daily flight utilization for the DC-10 fleet at nearly 9.6 hours; average DC-10 flight duration is 3.39 hours.

During 1986, the average distance flown by a DC-10 on each commercial flight was 1689 statute miles (2718 km). The shortest flight of 45 statute miles (72 km) was flown by United Airlines between Baltimore, Maryland and Washington, D.C. The longest nonstop revenue flight was 6995 statute miles (11,251 km) flown by Finnair between Helsinki, Finland and Tokyo, Japan. (*McDonnell Douglas Corporation.*)

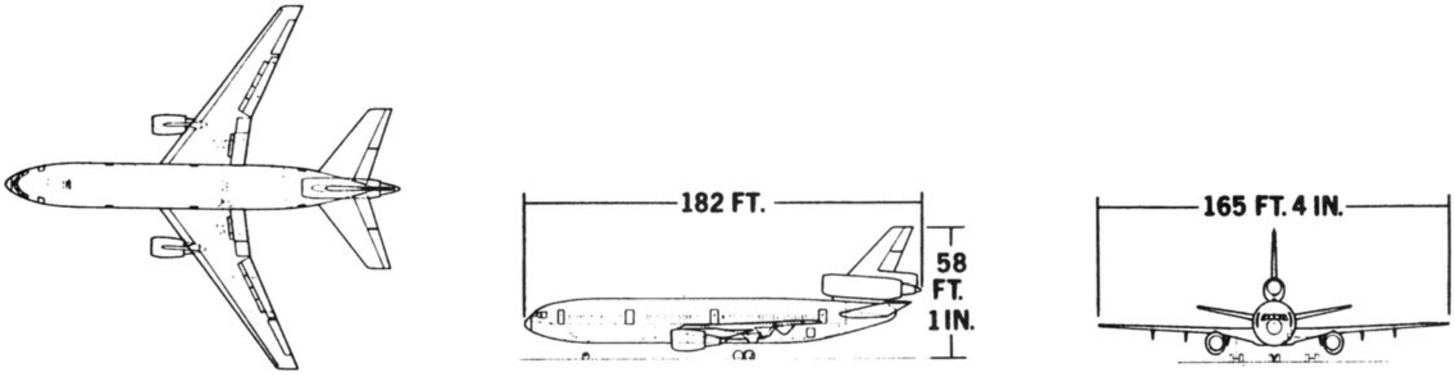


Fig. 2. General arrangement and overall dimensions of the DC-10 Series aircraft. (*McDonnell Douglas Corporation.*)

Fig. 3. Partially assembled DC-10 Series 30 aircraft (at height of production run in early 1980s), showing one of the three large jet engines in foreground being readied for attachment to the aircraft. (*McDonnell Douglas Corporation.*)

sonic space plane will be a commercial reality during the first quarter of the 21st century.

Some futurists, however, are already forecasting fundamental changes that may affect those commercial aircraft destined for introduction by the year 2025, plus or minus a few years. In terms of airplanes and associated aviation problems of that future period, the following points have been stressed:

1. Aircraft designers increasingly will be required to concentrate on sociological and market factors in addition to their past concern with the technical performance of an airplane. For example, with the grossly accelerating usage of telecommunications, the need for business travel will be curtailed. Telestaging of individual conferences and even of large group meetings will have a marked effect on business air travel, which has accounted for a substantial

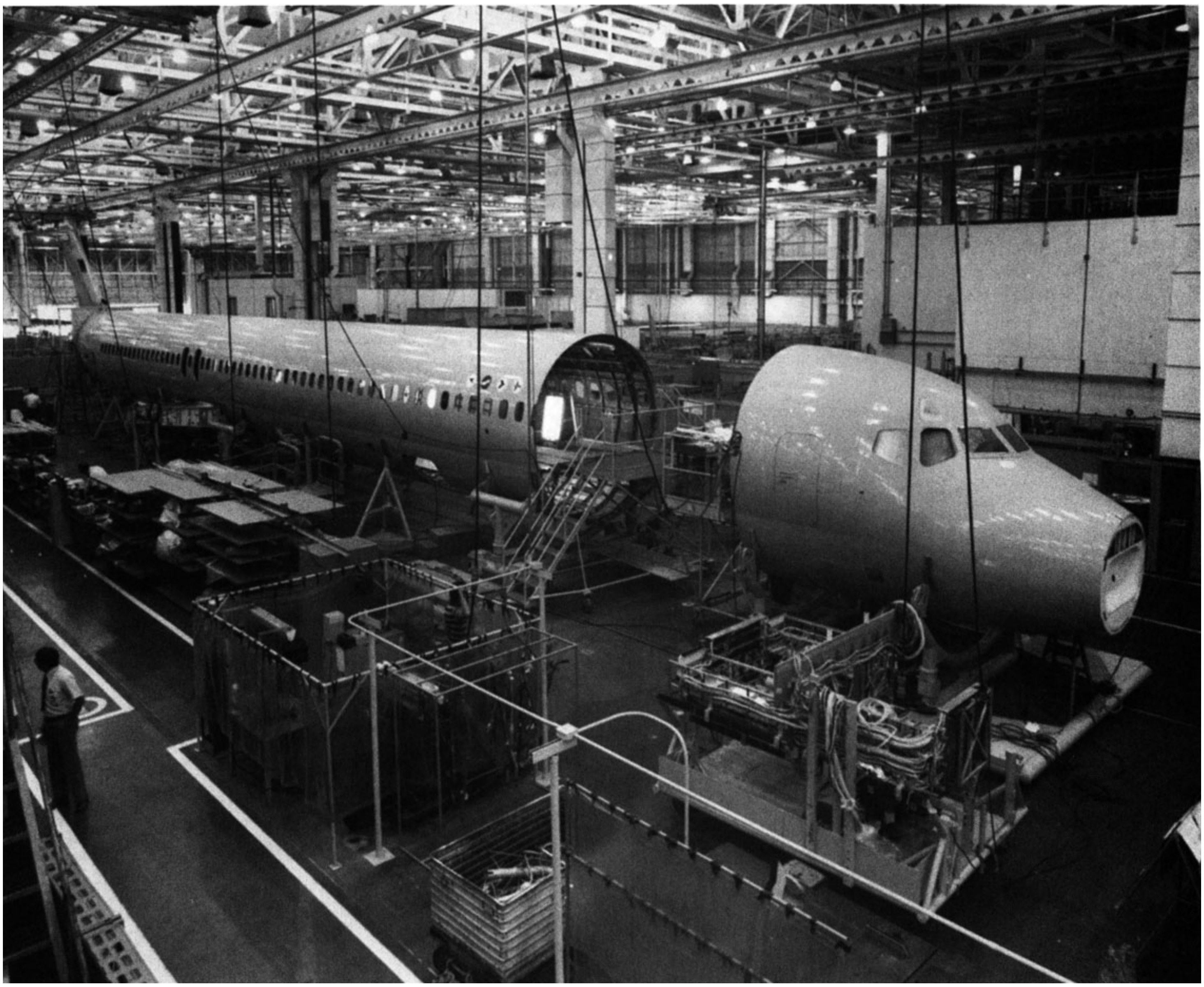


Fig. 4. Partially assembled DC-10 Series 20 aircraft. (*McDonnell Douglas Corporation.*)

percentage of airline revenues. Some futurists forecast that travel for pleasure and other personal and family reasons will comprise the major portion of airline revenues. This change could cause marked alterations in the interior design of an aircraft and include additional facilities for use during flight for attracting these travelers. With this class of passengers, aircraft speed may be deemphasized (flight duration lengthened). This could be translated into terms of what might be called "air cruises," which, in essence, would create new trade-offs for the aircraft designer. Futurists forecast that, during the second quarter of the 21st century, airliners with a capacity for 2000 passengers will be common. Flying at lower altitudes would be attractive to such passengers.

2. Some futurists forecast a growing demand for much larger cargo-carrying aircraft, citing the increasing acceptance of the "just-in-time" delivery of parts and components for consumption on manufacturing production lines. The existing large growth of air-delivered packages is expected to continue.
3. Forecasters also suggest that airplane designers must become more concerned with the total aviation system. Traditionally, the starting point in an air travel and cargo movement system has commenced with the airplane design per se, requiring that all other aspects of the system be geared to the aircraft. In essence, this spells out more attention to avionics. As pointed out by R. W. Simpson (Professor of Aeronautics and Astronautics, Massachusetts Institute of Technology), "Details on many factors, including wind and aircraft speed and direction, will probably be entered automatically into an enormous computer system that will route and reroute air traffic. Combined with a broad-area surveillance system that will likely rely on satellites instead of radar, improved computerization might even permit one or two controllers to handle thousands of flights at one time. Just a couple of people might direct all the air traffic over the North Atlantic."
4. In a systems approach, airplane designers, in addition to designing rapid and efficient passenger and cargo loading and unloading systems into the aircraft, may be called upon to extend their expertise to the total ground ingress and egress system.
5. Some futurists suggest that aircraft noise pollution remains as one of the principal unsolved problems. With the continuing growth of advocacy groups, this is considered a likely area of controversy of increasing magnitude and one that should be addressed more seriously by aircraft designers.
6. Although not directly within the province of the aircraft designer, there are in general numerous aviation problems of serious concern. One of the greatest, of course, is the need for improved air terminals. Considerable progress is occurring in air traffic control, ground systems, and land expansion, but futurists forecast that the situation in the year 2025 will be untenable in terms of capacity. Some experts suggest the construction of large artificial

> islands to serve as airports for urban complexes located on the seaboard or near inland lakes, such as Cleveland and Chicago. This was pioneered at least a half-century ago with the construction of the Lake Front Airport in Cleveland. However, such islands, filled deltas, and so on immediately conflict with some environmental and ecological concerns. See article on **Wetlands**.

Several additional futuristic concepts could be mentioned, but some tend to be of "whimsical uncertainty."

The 1990–1995 Airplane Design Era

The MD-11 (McDonnell Douglas) is designed with two fuselage lengths approximately 18.6 feet (5.66 m) longer than the DC-10 and an MD-11ER (extended range) that is approximately the same length as the DC-10. Three models include a passenger, an all-freighter, and a "combi," which carries passengers and freight on the same deck.

A wide variety of interior choices is available. Seating capacities on the standard airplane range from 276 persons in a three-class arrangement to 405 people in an all-economy configuration. Below deck, the MD-11 provides more combined containerized or pelletized cargo than available aircraft.

Maximum range is up to 7920 statute miles (12,746 km) nonstop, representing an increase of 24% as compared with the DC-10 Series 30 aircraft. A medium-range MD-11, derived from the standard aircraft, has a two-class capacity of 327 passengers and a nonstop range of 4590 miles (8080 km). The extended-range MD-11ER carries 277 first-class and economy passengers and has a nonstop range of 8870 miles (14,274 km), representing a 39% increase in range over the DC-10. See Figs. 5, 6, and 7.

In addition to the numerous engineering and technological features that contribute to the performance and safety of a passenger aircraft, operating economics also are a major consideration and, in essence,

Fig. 5. The MD-11 tri-jet. The design is based upon experience gained from over 25 years of improving the DC-10. The MD-11 features a smaller horizontal tail and winglets (vertical airfoils at the wing tips). Both of these features contribute to improved range by reducing in-flight aerodynamic drag. (*McDonnell Douglas Corporation.*)

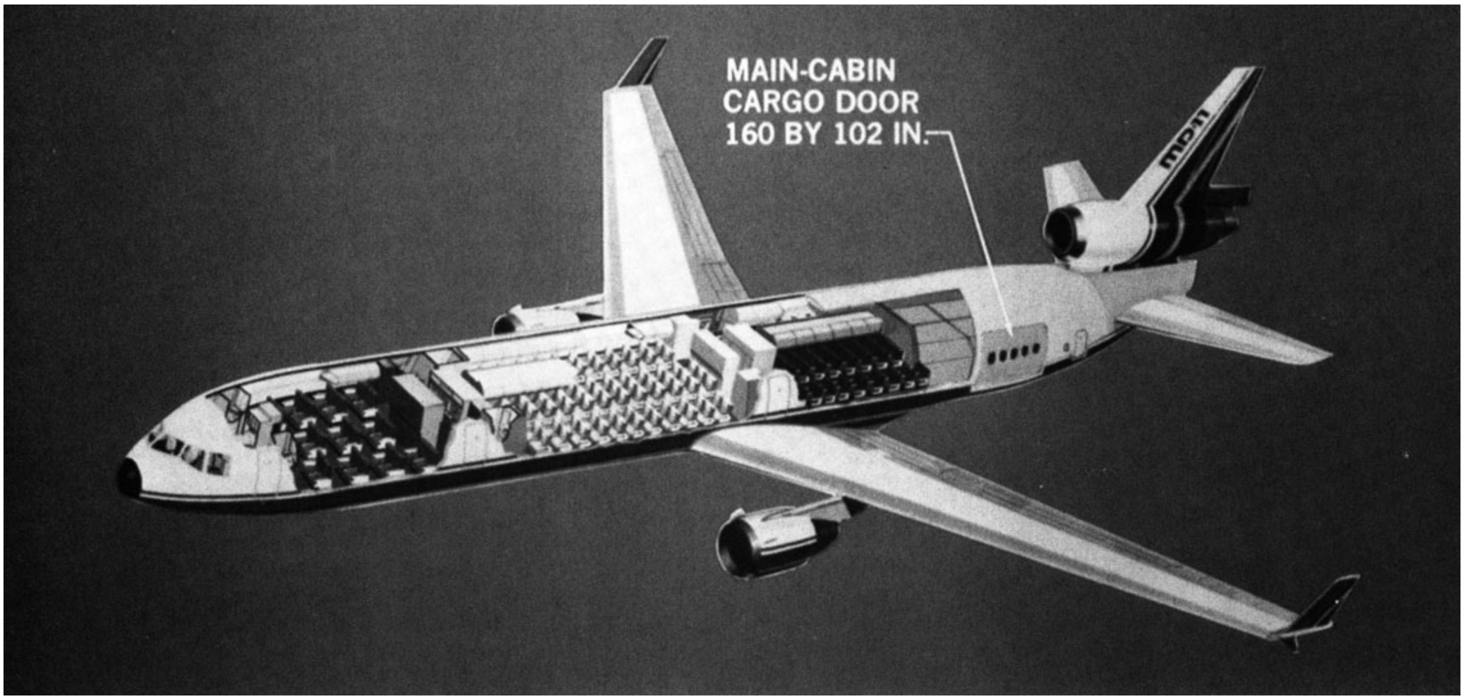


Fig. 6. The MD-11 COMBI provides a combination of passenger and cargo accommodations on the main deck. One design includes first-class and economy seating in the forward and mid-cabins for 214 persons, plus six pallets in the rear cabin. A large cargo door is provided in the aft fuselage. (*McDonnell-Douglas Corporation.*)

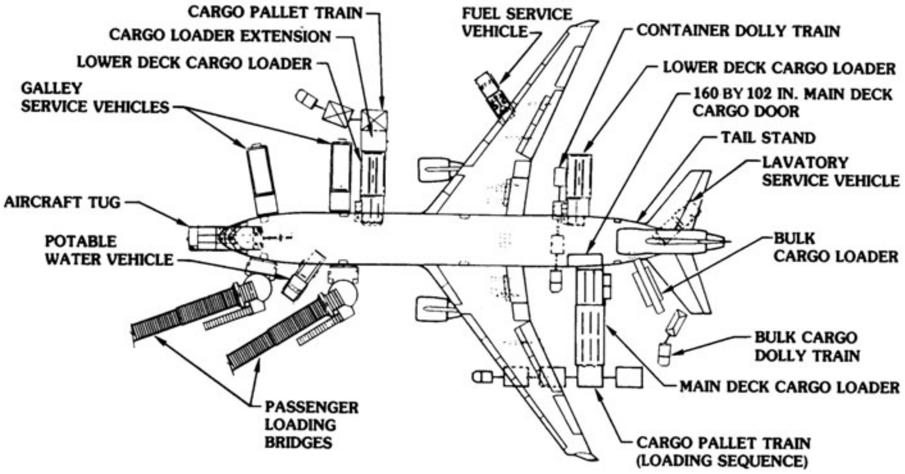


Fig. 7. Ground service equipment arrangement for the MD-11 COMBI. (*McDonnell Douglas Corporation.*)

determine preferences among the numerous airline purchasers of an aircraft. Looking to the late 1990s, the market requirements may be highlighted as follows:

1. *A growing need for new wide-cabin jets.* The availability of used wide-cabin aircraft has dropped by some 85% to a low level. By 1996, it is predicted that more than 500 of the world's wide-cabin fleet will be over 20 years old.
2. *More passengers, longer routes.* Through 1998, growth of more than 140% is projected for world passenger markets. More growth is seen for long-range direct services. The MD-11 will carry a mixed configuration of 320 passengers more than 6800 nautical miles (12,600 km).
3. *Greater cargo capacity.* This market is projected to grow more than 150% by 1998. Numerous older cargo jets are being retired. Consequently, in designing a new aircraft, such as the MD-11, substantially more below-deck revenue-cargo capacity must be designed into a wide-cabin passenger aircraft. This amounts to an increase of about 40% in the MD-11, as compared with the current DC-10.
4. *More interior flexibility.* The composition of multiclass service varies around the world and is constantly changing, currently in favor of a larger business-class configuration.
5. *Increased performance and profitability.* This requirement affects nearly all aspects of aircraft design and notably the size and power plant configuration. Other factors include more efficient usage of human power required to operate the aircraft and substituting, wherever practical, instrumental and computerized techniques for improving the efficiency of the crew.

In early 1969, Boeing introduced its "jumbo jet" airliner, the first of its 747 series. Over a period of approximately two decades, Boeing added eight different models, most having a gross weight of 836,000 pounds. These included the 747-200B, the 747-300, and a 747-Freighter model. The 747-100 was fitted especially to transport the space shuttle from California to NASA launching headquarters in Florida. The basic 747 aircraft was approximately 232 feet (70.7 m) long.

In the late 1980s, Boeing announced plans to build a 747-400 series of aircraft to enter service by the mid-1990s. The aircraft will feature the most advanced digitized avionics available.

Aircraft Instrumentation and Automation

In terms of *total systems* instrumentation, aircraft lagged behind the manufacturing and processing industries in the adaptation of instruments and automation. The very early planes of the pre–World War I era simply utilized a magnetic compass, a barometric altimeter, and an anemometer to indicate airspeed in the cockpit. Between the two World Wars, gyroscopes (vertical and horizontal attitude) and accelerometers (turning indicators) were added to the pilot's panel. Between the mid-1940s and mid-1960s, the pilot was aided by radio altimeters, weather radar, basic alarms for fuel, temperature, and landing-gear status, and analog airspeed and altitude indicators. Between 1965 and 1980, mechanized flight directors for guiding stick, rudder, and other control movements were added. Digital computers were used for the first time to monitor and indicate the status of the hydraulic and electrical systems. During the interim period between 1980 and the present, an array of electronic displays were added, as well as side-stick controllers, moving-map displays, collision avoidance systems, and flight management systems.

Automatic pilots for maintaining speed and direction were not added until the 1950s. During this same period, analog flight stability computers were first used. Automatic landing systems were added during the 1965–1980 period, as were hydraulic systems to assist in moving flight surfaces. Inertial navigation systems also appeared during this period.

Progress from 1980 to the 1990s included so-called "fly by wire" digital computers for actuating flight surfaces over an electric network. In special instances, additions included the global positioning navigation system and a microwave system for guiding an aircraft on a curved approach.

Shown schematically in Fig. 8 is the panel of controls planned for the forthcoming Boeing 747-400 series of aircraft.

Since the concept of an "automatic pilot" first was envisioned, there has been considerable disagreement among experts pertaining to the blending of automation with human factors engineering. Accidents have been attributed to the manner in which pilots have used (or misued) an airplane's automatic system. A system study of human factors and aircraft automation is now being undertaken by an interagency National Plan for Aviation Human Factors, coordinated by the Federal Aviation Administration. The results of these studies are expected by 1995.

Airplane Design Fundamentals

An airplane may be defined as an aerodynamically designed, winged, powered, heavier-than-air craft intended for sustained flight in the earth's atmosphere. A glider is an engineless airplane. The airplane is part of the total family of aircraft, which includes other heavier-than-air forms, such as helicopters and autogiros, and lighter-than-air forms, such as balloons and airships (blimps, dirigibles, zeppelins).

Various configurations of airplanes serve two major uses: (1) As vehicles for transporting people and goods; and (2) as military weapons systems. Other important uses, although relatively minor in terms of the number of aircraft used, include: (a) As carriers of airborne instruments for collecting geophysical, geodetic, and meteorological information (mapping, surveying, mineral and natural resource exploration, weather observation); (b) as searching tools—for lost persons and craft at sea, in jungles, mountainous terrain, and desert—and for determining the presence of dangerous conditions (forest fires, rising flood waters, hurricane centers, and movement); (c) as airborne inspection and

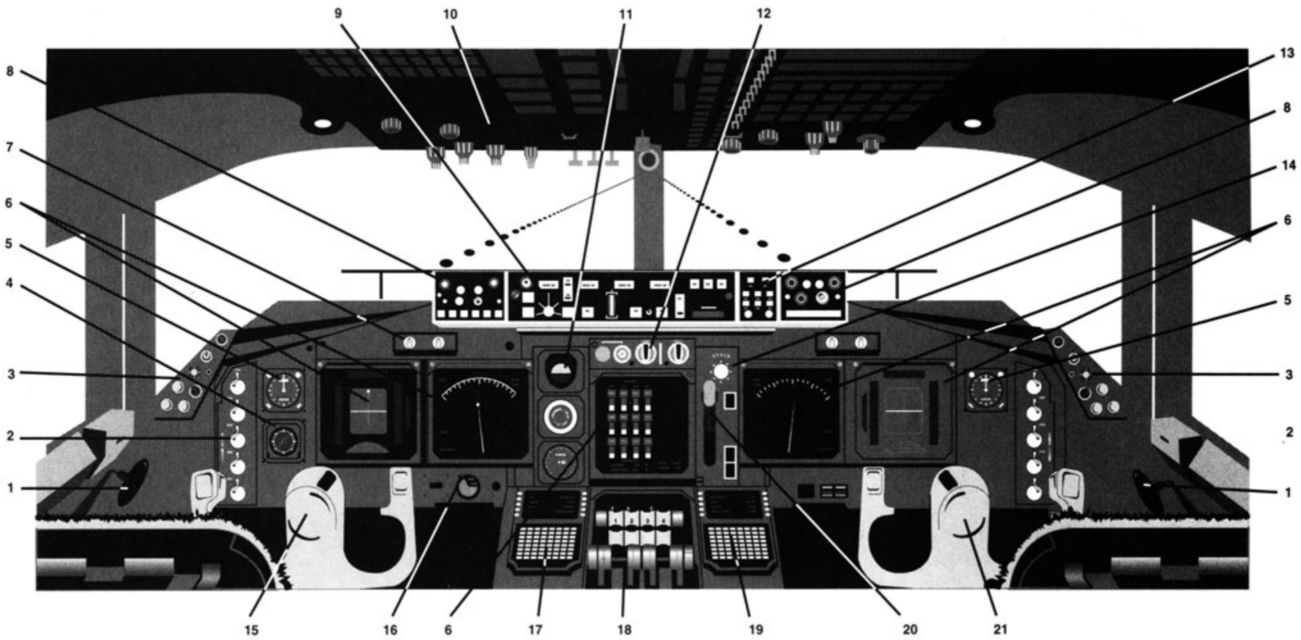


Fig. 8. Instrumentation and automation available to pilots in the Boeing 747-400: (1) nosewheel steering tiller; (2) instrument source selection; (3) internal lighting controls; (4) radio magnetic indicator; (5) real-time clock; (6) electronic flight instrument system; (7) display switching; (8) electronic flight instrument system control panel; (9) autopilot flight director system mode control panel; (10) overhead panel (engine start; electrical system; hydraulic system; oxygen system; fuel system; fire control; inertial navigation; system mode select; and temperature control); (11) standby instruments; (12) electronic flight instrument system control panel; (13) engine indication and crew alert system; (14) alternate flap controls; (15) captain's control column; (16) brake pressure indicator; (17) maintenance control display unit; (18) engine controls; (19) alternate gear controls; (20) landing gear extension handle; and (21) first officer's control column.

command posts for the supervision and management of geographically sprawling projects (ranching operations, construction projects, such as tunnels, highways, pipelines, commercial fishing operations)—however, these operations are frequently assigned to helicopters; (d) as "dusters" of crops for spreading insecticides and treating chemicals required by certain types of agricultural products; (e) as forest fire fighters, for carrying and dropping water and chemicals; and (f) as subsystems for controlling automotive vehicular traffic and for seeking out and helping in the apprehension of crime suspects—roles also often assigned to helicopters.

Military aircraft fall into three major classifications: (1) For bombing, strafing, and other offensive operations; (2) for fighting enemy aircraft in an effort to attain and maintain control over the air space of a given region—and to intercept enemy bombers and missiles; and (3) for conducting reconnaissance and command operations, including determination of enemy maneuvers and positions, as well as directing land and sea operations from an airborne observation platform. Most other military aircraft are used mainly for transporting people and equipment for numerous purposes.

There are, of course, still other specialized applications for commercial, military, and civil aircraft, not the least of which, in terms of small planes, is their use for recreation and the sport of flying.

With the perfection and refinement of helicopters a few decades ago, some of the former more specialized uses of aircraft have been assumed by the helicopter, with its superior maneuverability and hovering capability as, for example, in rescue missions.

While some valiant efforts were made in the 1800s and even before to obtain some form of sustained flight in a heavier-than-air craft, it is essentially universally accepted that this achievement was first successfully claimed by the Wright brothers of Dayton, Ohio. On December 17, 1903, the Wright brothers made several flights in their wood-and-muslin craft at Kill Devil Hills beach (North Carolina). The first successful flight had a duration of 12 seconds and a distance of 120 feet (36.6 meters). A later flight on the same day was of 59 seconds duration and covered a distance of 852 feet (259.7 meters). Less than two years later (October 5, 1905), in their third major design, the Wright brothers covered a distance of $24\frac{1}{5}$ miles (38.9 kilometers) in a time span of 38 minutes, 3 seconds, at a location called Huffman Prairie, near Dayton.

Although apparently very close to mastering powered flight, Prof. S. P. Langley of the Smithsonian Institution suffered two crashes in his *Aerodrome*, the last crash occurring only nine days prior to the successful flights of the Wright brothers. In the 1890s, the German engineer, Otto Lilienthal, had gained extensive experience with both monoplane and biplane gliders and was at work on motorizing one of his gliders. Unfortunately, he crashed in one of his gliders in 1896.

Propulsion

Following the success of the Wright brothers in developing the first true airplane power plant, engine development progressed rapidly. The Gnome "monosoupape" air-cooled engine of World War I fame was developed in France in 1909. This engine was unique in that the crankcase and cylinders rotated with the propeller about the stationary crankshaft. Good cylinder cooling was obtained by this arrangement, but engine lubrication was a difficult problem. Later in the war, this engine was producing 125 horsepower[1] at 900 revolutions per minute. Major developments by the British and American designers of that period were in liquid-cooled engines. Two famous engines of this type were the 8-cylinder LX 105, which produced 90 horsepower, and the 12-cylinder Liberty engine, which developed up to 400 horsepower. The latter engine was used in a number of different Allied combat airplanes during the latter part of the war and for several years thereafter. The 8-cylinder liquid-cooled Hispano-Suiza engine also powered many World War I airplanes and was the first to incorporate cast aluminum-alloy cylinder blocks with steel liners.

Advancements in aircraft propulsion diminished considerably for a number of years following the close of World War I. However, in the 1930s, the development of the controllable-pitch propeller was a tremendous advance and opened many new possibilities for airplanes of greater weight and speed. The next milestone was the perfection of the gasturbine (jet) engine into a useful airplane power plant. By the end of World War II, reciprocating engines were producing as much as 3500 horsepower and weighed less than one pound per horsepower. This was the peak of development for that type of engine. However, it could no

---

[1] One horsepower =1.014 metric horsepower.

longer compete with the gas-turbine engine which offered far more power output for each pound of installed weight.

Theory of Propulsion. A dynamic flying machine must be set in motion by a force called thrust, provided either by a propeller, the forward component of the lifting force of rotating helicopter wings, the thrust of gas-turbine engine exhaust, or the thrust of rocket engine exhaust. Gliders and soaring machines are initially set in motion by the thrust provided by a launching or towing device. Continued flight is maintained only by descending into rising air currents.

With an airplane, the thrust force must be equal to total airplane drag to maintain level flight at any given speed or altitude. Since airplane drag increases with speed, greater thrust is required for any increase in speed. Thrust force must also be greater than total drag when the airplane is accelerating, as at take-off, or when the airplane is climbing. Generally, an airplane has reached its maximum speed or altitude when all available thrust is being utilized. However, in some high-altitude jet airplanes, the amount of thrust available to meet take-off acceleration and high-altitude climbing requirements is somewhat greater than that required to obtain maximum permissible speeds at lower altitudes because of airframe structural limitations.

In the case of rocket-powered missiles, the thrust is utilized in a somewhat different manner. A large thrust force is used for a short, initial period to gain a very high speed. After the rocket motor has depleted its fuel, the missile continues on its trajectory by means of the energy imparted to it by the high initial thrust. Once outside the earth's atmosphere, space missiles and satellites will continue in motion indefinitely with no friction forces to absorb their energy, until pulled back again by the earth's gravitational field, or "captured" by some other large body in space.

Propellers. The function of a propeller is the conversion of engine shaft torque (turning force) into thrust. In the early days of propeller design, it was thought that thrust was obtained by the rear face of the propeller blade pushing against the air. Later, it was learned that propeller blades should be considered as small rotating airplane wings with the shape of the airfoil being of primary importance. Therefore, the upper surface of the airfoil, or the front rather than the back of the propeller blade, should be given primary design consideration.

Propellers may be installed in two different ways. In the "tractor" installation, the propeller is mounted forward of the wing. The tractor propeller "pulls" the airplane through the air and the thrust bearings in the engine are designed for thrust loads in this direction. Nearly all propeller-driven airplanes are designed for tractor-type installations because of the unrestricted air inflow to the propeller. As the terminology implies, "pusher-type" propellers are installed aft of the wing and push the airplane forward. The first Wright airplane was of the pusher type. The advantage of this arrangement in multiengined airplanes is that the absence of engine nacelles forward of the wing permits unrestricted airflow over the wing.

Two theories are recognized on the operation of a propeller in converting engine shaft torque to thrust. The momentum theory deals with the energy change to the air mass acted upon by the propeller. The other and most commonly used concept in determining propeller performance is the blade element theory. In this theory (known as the Drzewiecki theory), each propeller blade is considered as being composed of an infinite number of airfoils (called blade elements) joined end to end, forming a shape similar to a twisted airplane wing. Because the propeller is rotating about the engine shaft centerline, each blade element will be rotating in a different arc. The greater the distance of each blade element from the shaft centerline, the larger is the arc it circumscribes and thus the greater the distance it must travel for each revolution. Each blade element, therefore, is moving at a different velocity which is greatest at the blade tip. If each blade element is to operate at maximum L/D (lift/drag), the angle of attack α for each element must diminish as the distance from the hub increases. This gradual decrease in blade angle gives the propeller blade its twisted appearance.

Since the circumference of a circle is $2\pi r$, each blade element rotates through a circular path of $2\pi r$ during each revolution, with r the distance from the center of rotation, expressed in feet or meters. If the propeller is turning at n revolutions per second, the linear velocity of each blade element in its individual plane of rotation must be $2\pi rn$ feet (meters) per second. In flight, the propeller is moving forward at the same speed as the airplane, as well as rotating about its own axis. This forward velocity must also be expressed in feet (meters) per second to keep all terms of the equation in common dimensions. The total velocity of each blade element is the resultant of the two component velocities, the forward velocity V and the rotational velocity $2\pi rn$, or $2\pi rnV$. This is shown graphically in Fig. 9 for each blade element.

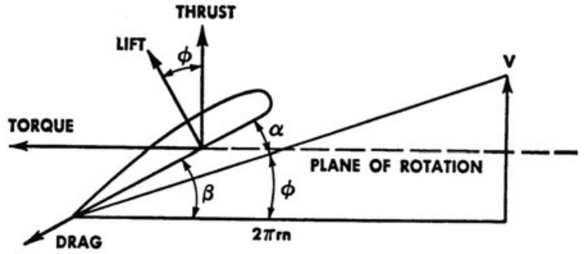


Fig. 9. Vector diagram of propeller blade element.

As shown on the diagram, the angle of attack α is equal to the fixed blade setting angle β, minus the angle ϕ which is determined by airplane velocity V and propeller rotational velocity $2\pi rn$ (or πdn since diameter $= 2r$).

The flow of air about the propeller blade creates lift and drag forces in the same manner as airflow about an airplane wing. The resultant of these forces can be divided into two components, the thrust component acting in the direction of the propeller axis and the torque component. The torque component acting in the plane of rotation resists the operation of the propeller and is the component that must be overcome by the engine shaft torque. Similar to the airplane wing, the magnitude of the thrust and torque components depends upon angle of attack and velocity.

The efficiency of a propeller in converting engine power to thrust is obviously of tremendous importance to airplane performance. The efficiency of a propeller during cruising flight determines the range of an airplane equally as much as the efficiency of the engine itself in converting fuel energy to shaft power. Propeller efficiency is expressed as the ratio of thrust power delivered to engine power required to turn the propeller. Thrust power is thrust force (pounds) multiplied by airplane velocity V (feet per second), or TV (foot-pound/second). The power required to turn the propeller may also be expressed in foot-pound/second by multiplying engine shaft horsepower by 550 since one horsepower is 550 foot-pounds/second.[2] The airspeed at which maximum efficiency occurs is called the design airspeed for the propeller.

Employing fixed-pitch propellers in high-performance airplane design presents numerous problems. If a wing airfoil section is chosen for a transport airplane that will result in maximum L/D at high cruising speeds, a propeller of relatively high angle must also be chosen to yield good efficiency at high cruising speed. However, this propeller will give poor high-speed performance as well as inferior performance at take-off and climb. A high-speed propeller chosen for a high-speed airplane will have reduced efficiency during cruising and give less than optimum range for the airplane, in addition to reduced climb and take-off performance.

The advent of the controllable-pitch propeller offered a solution to these difficulties. Provided with a means for varying pitch angle, the pilot could select the optimum angles for take-off, climb, efficient cruising, or high-speed flight. It is notable to observe that modern propeller airplane performance essentially commenced with the development of the controllable-pitch propeller. Fixed-pitch propellers are used only on small, light airplanes where cost and lightness of installation are major considerations. The speed range of such airplanes is limited by power output of the engine as well as the propeller.

The velocity of the blade elements has little effect on efficiency until velocities near the speed of sound are approached. When the velocity

---

[2] One metric horsepower =542.5 foot-pounds/second.

of the blade elements close to the tip approaches the speed of sound, the effect of compressibility increases power input requirements at no increase in thrust, thereby reducing efficiency. Since compressibility of an airfoil is affected by airfoil thickness ratio, propeller tips are built as thin as possible. Speed of sound is proportional to temperature and, normally, tip speed is limited to 800 to 1,000 feet per second (244 to 305 meters/second). To prevent tip speeds from exceeding these limitations, high-powered engines, requiring propellers of large diameter, are equipped with reduction gears. This arrangement permits lower propeller rpm than the engine rpm required for efficient engine power output.

In addition to providing efficient thrust during cruising conditions, the propeller must be able to convert total engine power to thrust for take-off. This is necessary to provide rapid acceleration to flying speed and keep take-off distances within reasonable limits.

Some important propeller terms include:

(*Fixed-pitch Propeller*)—A propeller with blade angle fixed.

(*Adjustable-pitch Propeller*)—A propeller on which blade angles can be adjusted on the ground with propellers not turning. This is especially convenient for a take-off on a hot day from a high-altitude airport.

(*Controllable-pitch Propeller*)—A propeller on which the blade angle may be changed in flight by controls in the cockpit.

(*Constant-speed Propeller*)—A controllable-pitch propeller with a speed governor which maintains selected rpm constant by automatically changing blade angle regardless of airspeed or engine power. Most propellers of this type also provide for full "feathering" of the blades in flight to prevent "autorotation" or "windmilling" of the propeller, thus reducing drag when the engine is dead.

(*Feathering*)—Mechanically increasing pitch angle until the blade is turned approximately to the direction of flight. Feathering is used most frequently on multiengine airplanes.

(*Reverse Thrust*)—Thrust in the direction opposite to flight. Nearly all multiengine airplanes utilize constant-speed, full-feathering, reversible-pitch propellers. The pitch change mechanism is designed to permit the pilot to select negative blade angles immediately after landing, thus creating reverse thrust for rapid airplane deceleration. This feature saves brakes on heavy airplanes and permits safe landings on icy runways where wheel braking is quite ineffective.

(*Dual Rotation*)—Two propellers on the same engine shaft and operating in opposite direction. One propeller shaft rotates inside the other while both are driven by reduction gears in the engine nose case.

(*Windmilling Drag*)—Perhaps the most undesirable action in the use of propellers as a thrust-producing device is the drag that results from propeller windmilling. If engine power is lost in flight, the airflow over the propeller blades will cause the propellers to rotate and, in turn, rotate the engine. When the blade angle is great enough to be commensurate with the flying speed, propeller efficiency will remain high and the drag will not be appreciably greater than engine load. However, if a failure occurs in the pitch-change mechanism and the blade angle is reduced, excessively high rotational speeds may result. The low blade angle and the resultant negative angle of attack of the blade airfoil element cause a tremendous reduction in propeller efficiency. This means the load required to turn the dead engine is supplied by an inefficient propeller, and the resulting windmilling drag can become dangerously high. In such instances, it is necessary to reduce airspeed or altitude to reduce propeller rotational speed. However, unless actual failure occurs to the pitch-change mechanism, the propeller speed governor will maintain rpm as it was before engine failure, and the feathering system will feather the propeller to stop windmilling rotation altogether. See also **Aerodynamics and Aerostatics; Helicopters and V/STOL Craft.**

Jet Propulsion. The jet or rocket engine moves the airplane or missile forward by reacting to the momentum of a mass of air or gaseous matter accelerating out the rear (nozzle) of the engine. The recoil of a rifle and the reactive thrust of a fire hose are typical examples of this principle. The thrust force is not the result of the jet gases pushing against the air behind, but instead is the reaction to the momentum of the escaping gases. This fact is demonstrated by rocket motors, which produce full thrust in outer space beyond the atmosphere.

Power Measurement. Jet thrust force is measured in pounds as is the thrust force of a propeller. The magnitude of jet gross thrust force is determined by the mass flow of air or gaseous matter multiplied by the velocity with which it escapes from the nozzle. This may be expressed by

$$F = MV_j \tag{1}$$

where F = thrust force, pounds
M = mass flow rate = W/g
V_j = jet velocity, feet per second
W = weight flow, pounds/second
g = gravitational acceleration + 32.2 feet/(second)^2 [9.8 meters/(second)^2]

Because aerodynamic equations require thrust to be expressed as a force, to equate thrust requirements against drag and acceleration forces, there usually is no reason to express jet engine output in terms of horsepower. But, where the equivalent is desired, jet engine "power" can be determined by multiplying thrust force times flying speed in feet per second and dividing by the horsepower constant.

$$\text{Horsepower} = \frac{(\text{Jet Thrust}) \times (\text{True Airspeed}) \times 1.467}{550} \tag{2}$$

Therefore, when flying speed is 550 feet per second (375 miles (603 kilometers) per hour), the jet thrust and jet "power" are the same. At speeds below this value, jet "power" is less than jet thrust in pounds. In the case of the turboprop engine, where both propeller thrust and jet thrust are utilized, it is sometimes desirable to convert jet thrust into horsepower. By adding this value to the horsepower rating of the engine, the total propulsive output of the engine can be expressed as "equivalent shaft horsepower" (eshp). Jet horsepower must be multiplied by propeller efficiency to convert to shaft horsepower. Assuming a propeller efficiency of 80%.

$$\text{eshp} = \frac{(\text{Jet Thrust}) \times (\text{True Airspeed}) \times 1.467 \times 0.80}{550} + \text{Shaft Horsepower} \tag{3}$$

Propulsive Efficiency. The efficiency with which a propulsive system is utilized in propelling the airplane or missile is called propulsive efficiency. A thorough understanding and investigation of propulsive efficiency are necessary for successful design when comparing one propulsion system with another. This is especially true when comparing propeller propulsion with jet propulsion.

Propulsive efficiency may be expressed by

$$\text{Efficiency} = 2V_a/(V_J + V_a) \tag{4}$$

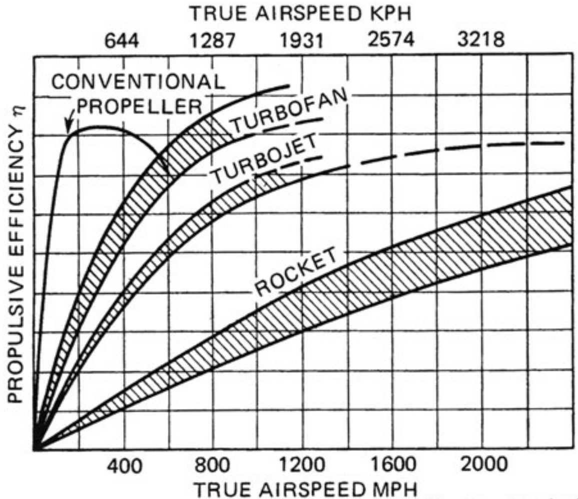
where V_a = airplane velocity, feet per second
V_J = jet velocity (or propeller slipstream velocity), feet per second

In the case of propeller-driven airplanes, propulsive efficiency very closely approximates propeller efficiency because the propeller moves a large mass of air at relatively slow speeds. During cruising conditions, the air behind the propeller is moving at approximately the same speed as the airplane; hence, V_a and V_J are approximately equal. In contrast, the thrust-producing gases escaping from the jet engine nozzle are extremely high in velocity, relatively smaller in mass flow, and cannot be varied extensively in flight.

Because it is difficult to design a pure jet engine that would have satisfactory fuel economy and light weight without high jet velocity, jet-powered airplanes are generally designed for high-speed flight to obtain maximum propulsive efficiency.

High-altitude operation must be considered in propulsive efficiency analysis. Because of compressibility effects on the propeller airfoil, propeller efficiency decreases with both airspeed and altitude. This factor limits propellers to approximately 500 to 600 miles (805 to 965 kilometers) per hour true airspeed and 40,000 to 45,000 feet (12.192 to 13.716 meters) altitude. In contrast, since jet thrust is accomplished at high jet velocities, the higher the airplane speed, the greater is the propulsive efficiency. A comparison of propulsive efficiency for the various propulsion methods is given in Fig. 10.

Turbojet Engines. The principle of jet propulsion was recognized early in aviation development as possibly the best method for obtaining greater speeds. The practical application of the principle proved



Hamilton Standard

Fig. 10. Comparison of propulsive efficiency.

difficult, however, because of the efficiency of early engine designs. Early developments centered around the turbojet and turboprop engines, patterned to some extent upon developments in stationary gas-turbine and steam-turbine engines. In 1930, Frank A. Whittle was granted an English patent for a turbojet engine design, but it was not until 1937 that intensive work was commenced on his design in England. The Germans initiated turbojet development in 1936 along with other jet engine and rocket motor programs. This work was carried on in an intensified manner, and the first turbojet-powered airplane flight took place in Germany in 1939. The airplane was a Heinkel HE178 powered by a Heinkel HE53B engine. Meanwhile, the Whittle engine program progressed rapidly in England. In 1941, the fourth Whittle engine, Model W1 delivering 855 pounds of thrust, propelled a Gloster E-28 airplane to 339 miles per hour at 20,000 feet. In this same year, the General Electric Company agreed to develop turbojet engines for the U.S. Army Air Forces from plans of the Whittle W2-B engine. The first American flight took place in 1942 when a Bell Aircraft P-59, powered by two General Electric IA engines of 1,300 pounds of thrust each, flew successfully. By 1944, production was started on the 4,000-pound thrust J-33 engine for the P-80 airplane. During 1944, the first axial-flow compressor turbojet engine was successfully tested by General Electric.

The three basic parts of the turbojet engine are: (1) air compressor; (2) combustion chambers; and (3) a turbine wheel(s) for driving the compressor. The compressor in the forward part of the engine draws in large quantities of air, compresses it to high pressure, and forces it through the combustion chambers. Fuel is forced through spray nozzles into the combustion chambers where it is mixed with part of the air. This mixture is then burned in a continuous combustion process and produces a very high temperature (4000°F; ~2205°C or higher), which heats the entire air mass to 1600 to 2000°F (871 to 1903°C) or higher. The hot gases then pass through the turbine wheel(s) directly behind the combustion chambers. Part of the energy in the hot gases is used to rotate the turbine wheel, which in turn drives the compressor by means of a direct shaft. The hot gases then expand and blast out of the tailpipe nozzle at high velocity. The reaction to the momentum of the escaping gases is the thrust force which propels the airplane.

The efficiency with which the three basic parts of the turbojet engine operate, particularly the turbine wheel(s) and compressor, is very important to thrust output. Since approximately two-thirds of the energy available in the hot gases is absorbed by the turbine wheel to drive the compressor, the need for high efficiency is obvious. Early models failed because of inefficient turbine and compressor design, coupled with inferior materials of construction. Since thermodynamic laws reveal that the energy available in hot gases increases with both temperature and pressure, design development constantly strives for compressors that will provide even greater pressures and turbine wheels that will withstand higher temperatures. One of the most outstanding improvements made has been that of more effective materials for turbine-wheel blades.

Special nozzles are used to reduce the disturbing noise intensity of the jet blast during take-off. This low-frequency, high-intensity noise is caused by the turbulent mixing of jet exhaust gases with the atmosphere. Nozzles reduce the formation of large-scale eddies, lowering the intensity of the lower "rumbling" frequencies and changing the character of the sound. Also, by permitting a more rapid mixing of the jet gases with surrounding air, they increase the rate of sound-energy dissipation.

Thrust reversal to assist wheel braking during the landing roll is accomplished by stopping the aft flow of the jet and directing it partially forward in the direction of airplane movement. To prevent the hot gases from entering the compressor inlet, of the same or adjacent engines, the reverser directs the jet blast outward at a suitable angle from the nacelle. Because of this angle, and losses due to turning the path of the gases, thrust available in the reverse direction is about 40% of forward thrust for the same engine rpm.

The magnitude of thrust output of a turbojet engine is determined by the mass airflow through the engine and the velocity of the mass leaving the jet nozzle. This is expressed by previously given equation (1). That equation represents the thrust output when the engine and airplane are at rest. When the airplane and engine are in motion, air is being rammed into the engine compressor inlet, tending to retard the forward motion. This retarding force is called *ram drag* and its magnitude is determined by the mass air flow through the engine and the airplane velocity:

$$\text{Ram Drag} = MV_a \tag{5}$$

The net thrust for the engine in motion is, therefore, the *gross* thrust output minus the ram drag:

$$F = MV_J - MV_a \tag{6}$$

or

$$F = M(V_J - V_a) \tag{7}$$

Thus, the net thrust is equal to the product of mass airflow and the change in velocity through the engine. Actually, the mass flow of gases from the nozzle is increased by the amount of fuel added for combustion. However, since the airflow is about seventy times greater than fuel flow, the mass of the added fuel is neglected in basic equations.

Augmentation. This term applies to means for increasing thrust during take-off in addition to that delivered by the basic engine. *Afterburning* implies that fuel is added to the hot gases in the tailpipe after they have passed through the turbine wheel. The burning of this additional fuel greatly increases the gas temperature and correspondingly the jet velocity, resulting in as much as 50% increase in thrust. Because the hot gases pass through the tailpipe at high velocity, it is necessary to introduce the fuel through flameholders in the tailpipe similar in some respects to the flameholders used in the combustion burners. The increase in gas temperature from afterburning causes the gases to expand, which tends to increase tailpipe pressure and turbine wheel temperatures. *Water injection* can produce about 25% additional thrust. A desirable feature of this method is the small increase in engine weight over the basic engine as compared with considerable weight increase for afterburning. Water injection equipment causes little change in normal engine performance, whereas afterburning causes a slight increase in fuel flow during normal (afterburner off) operation because of the blocking effect of the flameholders in the tailpipe. Water is introduced into the compressor or is forced through nozzles into the combustion chamber, or both. When introduced into the compressor, the water cools the air by evaporation, permitting the compressor to deliver more mass flow of air at maximum rpm. To offset the additional airflow, a signal is sent to the fuel control, increasing fuel flow slightly to heat the additional air. Thrust is increased by the increased mass airflow, plus the mass of the added water and fuel. When water is introduced into the combustion chambers (or diffuser section), the resulting increase in compressor airflow is less than by direct compressor injection. However, large amounts of water injected into the combustion chambers, plus the slight increase in compressor airflow, result in a substantial gain in thrust. In the combustion-chamber injection method, alcohol may be mixed with the water to assist in heating the added mass rather than increasing fuel flow. The injection water must be very pure to prevent mineral deposits in the machinery.

Turbojet engines can be classified by the type of compressor used. The *centrifugal compressor* was used in early engines and in some engines today. In the centrifugal compressor, air enters the rotating impeller at the center through the inlet guide vanes. Rotating at high speed, the impeller compresses the air by centrifugal action. The air passes from the rim of the impeller through a diffuser, which retards the tendency of the air to rotate with the impeller, into the combustion chambers. The *axial-flow compressor* is used on large turbojet and turboprop engines. An advantage is the small diameter of the compressor. Many stages can be added together without increasing the diameter of the engine. See also **Gas and Expansion Turbines.**

Turbofan Engines. These engines were developed to provide greater propulsive efficiency than obtainable with conventional turbojets. This is accomplished by increasing the total compressor airflow and reducing the net velocity of the jet. The additional airflow does not pass through the combustion chambers and turbine wheels, but is directed aft by way of a fan outside the engine cases. In one engine design, about 25% of the air from the forward compressor is ducted outside the engine case and is mixed with the hot gases before passing through the nozzle. This type of engine is called "bypass" by the British. Bypass ratios permit from 30 to 60% or more of the propulsive force to be produced by the fan. In comparison, turboprop engine designs generally permit an energy distribution of 90% for the propeller and 10% for the jet nozzle. A basic difference between the turbofan and the turboprop is that the airflow through the turbofan is controlled by design so that the air velocity relative to the fan blades is unaffected by airspeed of the airplane. This eliminates the efficiency loss at high airspeeds which limits airspeed capability of the turboprop engine.

Turboprop Engines. Turboprop engine development began in 1926 when the English scientist Griffith proposed an axial-flow compressor and turbine engine for driving a propeller. After wind tunnel tests, development was discontinued because of the economic depression and scientific reluctance to accept the engine as having any practical significance. Following the successful development of the turbojet engine, interest was again directed toward the turboprop engine and successful flights in England and the United States were made with turboprop installations in 1945. For several years following these flights, the majority of effort was devoted to improving turbojet design in the United States and turboprop development lagged. More emphasis was placed on turboprop development in England and France during this period and the turboprop powered English Viscount 630 was placed in commercial passenger service in 1950.

Basically, the turboprop engine is a turbojet with most of the heat energy converted to shaft power by the turbine wheels, leaving very little for jet reaction. The basic parts of the engine compressor, combustion chambers, and turbine wheels are the same in both engines, except that larger or a greater number of turbine wheels are used. In most turboprop engines, the compressor rotor shaft extends forward and drives the propeller shaft through a system of reduction gears.

A significant difference exists between turboprop and turbojet engine controls. In the turbojet engine, fuel is metered by the control as required in proportion to airflow to maintain constant rpm, or is metered in the proper amount to effect changes from one rpm to another. The control must be designed to meter fuel within precise limitations to prevent overheating of the turbine wheels, compressor surge, or burner blowout. The control must also correct for varying conditions of altitude, air temperature, and airspeed for a given power lever position.

All of the control requirements of the turbojet are present in the turboprop engine with additional requirement imposed by the propeller. By means of the blade pitch-angle control, the speed and power absorption of the propeller can be varied independently of the normal engine control. In addition, the rather large inertia (resistance to rapid rpm change) of the propeller and direct-connected engine rotor further complicates the control requirements. To provide the pilot with a simple single-lever control, as used for the turbojet, the propeller control of the turboprop is integrated with the engine fuel control. With most turboprop controls, the engine is operated within a narrow range in the high-rpm region for all flight conditions. This operation minimizes the effect of propeller inertia on power changes.

Ramjet and Pulsejet Engines. The ramjet or athodyd (aerodynamic-thermodynamic-duct) was invented by Lorin in France in 1913. The pulsejet, or resonant-type jet propulsion motor, was invented in 1910, also in France. The only extensive use of the pulsejet engine was made by the Germans during World War II when it was used to propel the V-1 "buzz bomb." Intensive development of the ramjet engine principle was begun in the United States following World War II for use with long-range, high-altitude guided missiles.

Pulsejet Engines. Because of its inefficiency and principle of operation, the pulsejet engine cannot compete with the turbojet or ramjet engines and finds no practical applications. With reference to Fig. 11, the pulsejet engine consists simply of a combustion chamber with an air intake, incorporating check valves at the forward end, and a nozzle for expelling the gases at the aft end. Forward velocity is required in order to ram air into the combustion chamber to start the combustion process. With air pressure in the combustion chamber, fuel is added and the mixture is ignited by a spark plug. The resulting pressure in the chamber closes the check valves in the air inlet and forces the hot gases out of the nozzle. The velocity energy of the gases passing out of the nozzle reduces the pressure in the combustion chamber below the atmospheric level, causing the air inlet valves to open and admit a fresh charge of air. The combustion process is then repeated and, after a few cycles, is self-sustaining and requires no further spark ignition. For the engine to operate properly, it is necessary that the vibratory resonance of the air inlet valves be "tuned" with the impulses of the gases and sound waves escaping from the nozzle. This is accomplished by shortening or lengthening the tailpipe until resonance is established for the desired flight velocity. The V-1 engine used a grill of thin metal reeds, similar to reeds of a harmonica, for the air-inlet valves. The resonant frequency of explosions for the V-1 engine was approximately 200 to 300 Hz, creating the buzzing sound for which the missile was named.

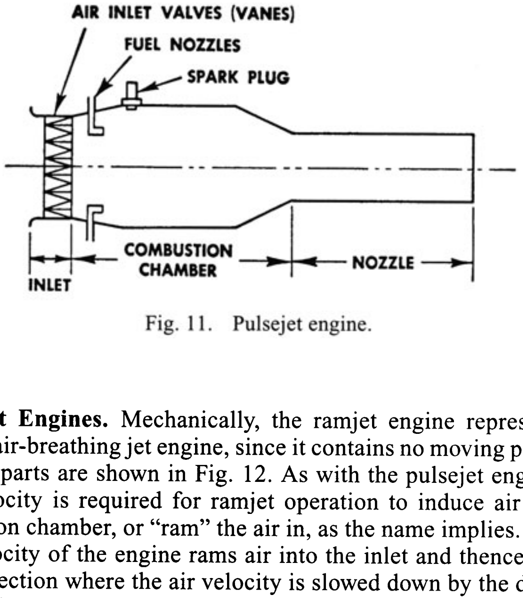


Fig. 11. Pulsejet engine.

Ramjet Engines. Mechanically, the ramjet engine represents the simplest air-breathing jet engine, since it contains no moving parts. The principal parts are shown in Fig. 12. As with the pulsejet engine, forward velocity is required for ramjet operation to induce air into the combustion chamber, or "ram" the air in, as the name implies. The forward velocity of the engine rams air into the inlet and thence into the diffuser section where the air velocity is slowed down by the divergent walls of the diffuser. In the combustion chamber, heat is added by the combustion of kerosine or other jet engine fuels which tends to expand the gases and increase their pressure. Because the entering ram air and the diffuser prevent expansion of the hot gases in a forward direction, the gases accelerate out the nozzle at a higher velocity than the velocity of the entering air. The reaction to the rearward velocity of the gases produces forward thrust.

The diffuser section is important in the functional operation of the engine. The outward taper must be carefully designed to permit proper slowing of the ram air with a minimum of losses. The shape of the air inlet is equally important, as is the design of the nozzle. In ramjet engines, combustion is continous after having been started by spark ignition or other methods. Because of the rather high velocity in the combustion chamber (300 to 500 feet (91 to 152 meters) per second),

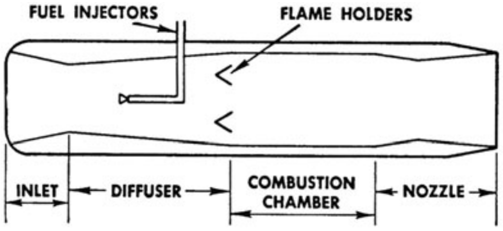


Fig. 12. Ramjet engine.

flameholders are required, as in the combustion chambers of turbojet engines.

For long-range missle applications, provided the missle remains at an altitude where there is sufficient oxygen, the ramjet engine has the advantages of light weight and no requirement for oxidizing chemicals (also keeping the weight down). These advantages also spell out low initial cost.

Modern Jet Engines. Sir Frank Whittle (United Kingdom) and Hans J. P. von Ohain (Germany) independently invented jet engines with improved features. Development of jet-powered military aircraft commenced during the late phases of World War II. During that period, the Germans developed the fastest jet-powered military aircraft—a plane that achieved a speed of about 525 mph. However, the lifetime of the aircraft was about 25 hours. This jet engine life was cost prohibitive. By comparison, modern jet engines have a lifetime ranging from 10,000 to 20,000 hours. The early jet engines developed a thrust of from 1000 to 2000 pounds. Modern jet engines develop up to 80,000 pounds of thrust. Pressures created in the combustion chambers of modern jet engines range up to 35 times that of the outside air. Combustion chamber temperatures now range up to 2900°F (1503°C) as compared with 1500°F (815°C) of the early jet engines. Fuel efficiency has increased by approximately a factor of 3 as compared with earlier piston engines. These performance increases have stemmed from the engine's internal design and the development and selection of superior materials.

In 1991, Sir Frank Whittle (now adjunct research professor at the U.S. Naval Academy, Annapolis, Maryland) and J. P. von Ohain (Senior Research Engineer, University of Dayton Research Institute, Dayton, Ohio) were jointly awarded the esteemed Charles Stark Draper Award for outstanding achievements in engineering. Of course, the engineering staffs of General Electric and Pratt and Whitney, among other specializing firms, also contributed to the modern jet engine for aircraft.

The engine for the new McDonnell Douglas MD-11 series of airplanes, previously mentioned, is a high–bypass ratio turbofan engine, ranging from 48,000 to 60,000 pounds (213.5 to 267 kN) of takeoff thrust. The PW4000 (Pratt and Whitney) engine also is suitable for use on current wide-body aircraft. The engine, shown in Fig. 13, is designed to provide 7% better fuel economy for the same size engine. Also, the innovative design has 50% fewer parts, which translates into fewer spare parts inventorying problems and less maintenance time.

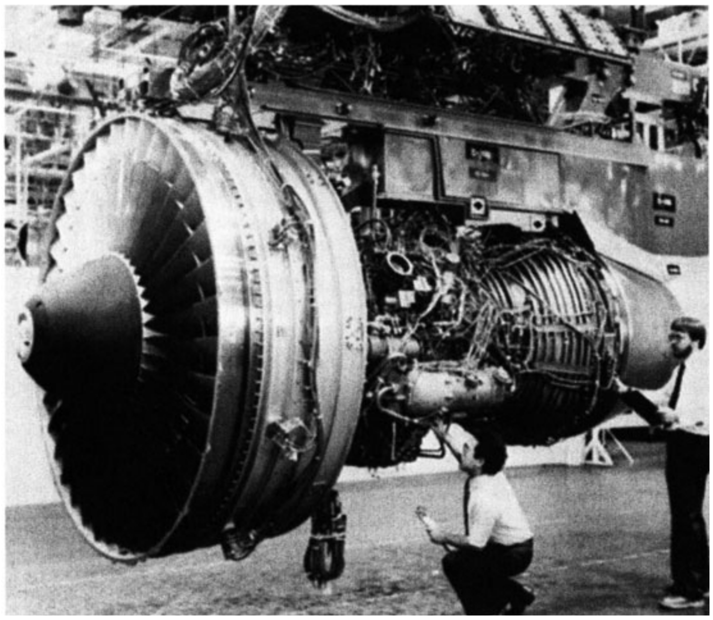


Fig. 13. High–bypass ratio turbofan jet engine of type designed for the McDonnell Douglas MD-11 aircraft. (*United Technologies, Pratt and Whitney.*)

Historical Aircraft

Early History. The following paragraphs are included here for those readers who may be interested in the early days of aircraft. Difficult to comprehend for all but the older readers, the early days of flying were indeed thrilling times and the phrase "air minded" was very popular in that era preceding the late 1930s. There was much competition in terms of aircraft performance and pilots. Unquestionably, the major meets and well-funded competitions, such as the Gordon Bennett Cup, the Schneider Cup, the Pulitzer Trophy, the Thompson Trophy, and the Bendix Trophy greatly catalyzed the development of superior airplanes. The progress in airplane design also was grossly assisted during the two major World Wars. It is sometimes difficult to realize, in perspective, that modern massive movements of people and cargo by air all have occurred within the life span of many people still living. Progress made during the early period of 1906 to 1914 is shown in Fig. 14. The speed records of the major competitions held during the period 1909 to 1939 are plotted in Fig. 15. The Reims (France) Meet of 1909 was the first extensive meet in aircraft history. The Gordon Bennett meets, commenced in 1909, were the first airplane speed races. The Schneider Cup meets were designed to promote the advancement of seaplanes and flying-boats and were confined to that category of aircraft. A primary incentive of the Pulitzer Trophy races was that of promoting the design of more effective fighter aircraft for the military. These meets were abandoned when U.S. General Mitchell was court-martialled. The Thompson Trophy races were sponsored by Cleveland industrialist, Charles E. Thompson, and were geared to speed. The Bendix Trophy races were of a different nature, namely, cross-country flights from Burbank, California to Cleveland, Ohio, and thus the speeds represented an average over long straight distances. These races highlighted the role of women as competent pilots, the 1938 race having been won by Jacqueline Cochran.

There was essentially no aircraft industry at the time of World War I in the United States, with just a little over 50 planes produced in 1916. The war, however, hastened design and production activities even though few American planes were used during the war. American pilots mainly flew French aircraft. The French introduced nearly 20 new planes during the period between 1914 and 1917. The fastest of these was the Spad S XIII, with a speed of 130 miles (209 kilometers) per hour and a ceiling of 22,300 feet (6797 meters). It had a 200-horsepower Hispano-Suiza engine and a wingspan of just a little over 26 feet (7.9 meters). The British also introduced nearly 20 planes during the same period. The S.E.5a was the fastest plane of that period, reaching a top speed of 138 miles (222 kilometers) per hour. It was powered by a 200-horsepower Wolseley Viper engine, had a wingspan of nearly 27 feet (8.2 meters), a range of about 340 miles (547 kilometers), and a ceiling of 19,500 feet (5944 meters).

Although Italy only introduced seven aircraft during the 1915–1917 period, it was the first country to employ aircraft in warfare. The fastest of the Italian planes was the Ansaldo S.V.A-5, with a speed of 143 miles (230 kilometers) per hour, powered by a 220-horsepower S.P.A. 6A engine, a wingspan of 32 feet (9.8 meters), and a ceiling of 22,000 feet (6706 meters). The fastest plane produced by the Germans was the Hansa-Brandenburg D1, with a speed of 116 miles (187 kilometers) per hour, a wingspan of a little over 26 feet (7.9 meters), powered by a 160-horsepower Austro-Daimler engine. It had a ceiling of 16,400 feet (4999 meters). Probably the most famous of the German World War I planes were those in the Albatros series, most of which were powered by Mercedes engines. The first airplane to have a machine gun synchronized with the propeller was the Fokker E-III, introduced in 1915. Other famous names in production in Germany during the war period included the Etrich A-11 Taube, introduced in 1910 and extensively used for pilot training; the Fokker E-III and D-VII; the LVG C-1, the Rum per C-1; the Lloyd C-II (Austrian); the Pfalz D-III; the Hannover CL-

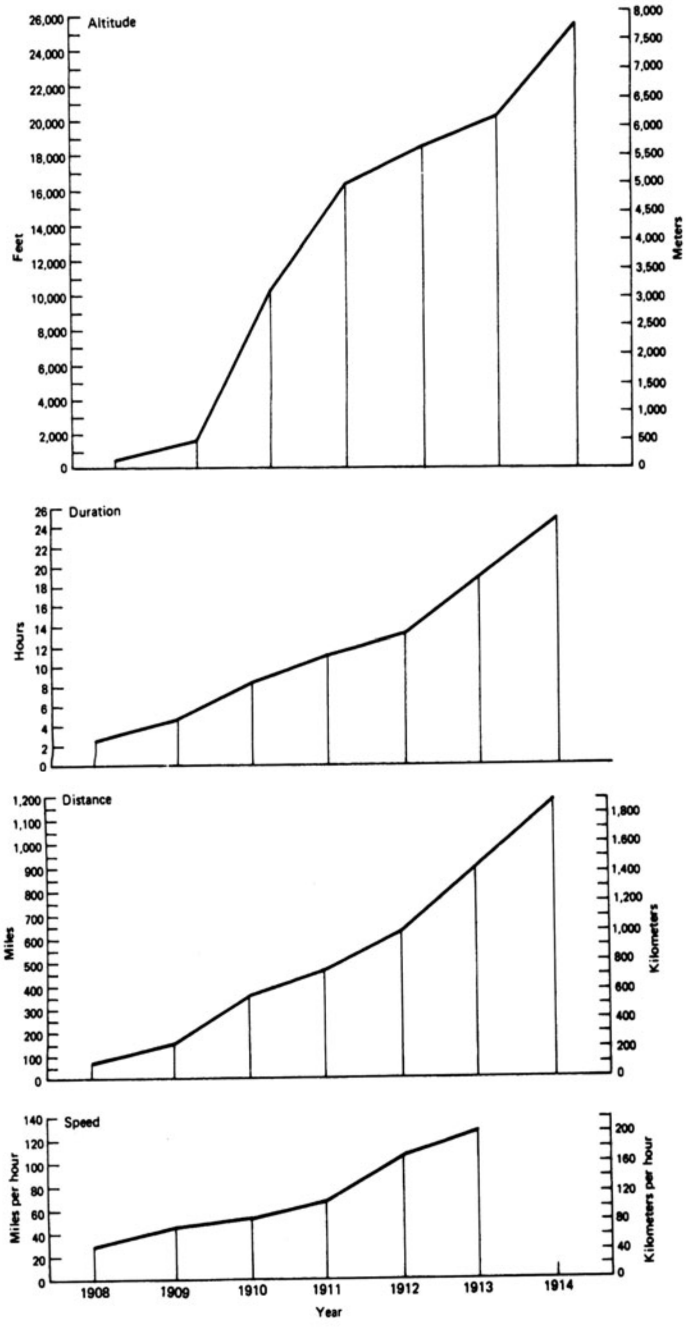


Fig. 14. Records set by early airplanes during period (1908–1914).

IIIa; the Gotah G-VIII; the Friedrichshafen G-III (a heavy bomber). The Zeppelin R-VI (not to be confused with a dirigible) was also introduced during that period.

Airplanes of World War II. Airplane production during the period just prior and carrying through the war years numbered in the tens of thousands, notably in the United States, Great Britain, Germany, Russia, and Japan. It is interesting to note that concentration on production of conventional propeller airplanes essentially delayed final development and refinement of jet aircraft during that period. The only true jet aircraft to see action during World War II were the British-built Gloster Meteor III and the German-built Me-262 jet fighters. The first mission for the Meteors, of which 280 were produced, was set for mid-April 1945 for the purpose of intercepting the German jet fighters. The encounter did not occur, however, and the war was ended without the Meteors seeing action. The meteors were equipped with two jet engines, each capable of 2000 pounds (907 kilograms) of thrust. They had a speed of 493 miles (793 kilometers) per hour, a range of 1340 miles (2156 kilometers) and a ceiling of 44,000 feet (13,411 meters)—with a crew of one. The German Messerschmitt Me-262A Sturmvogel was

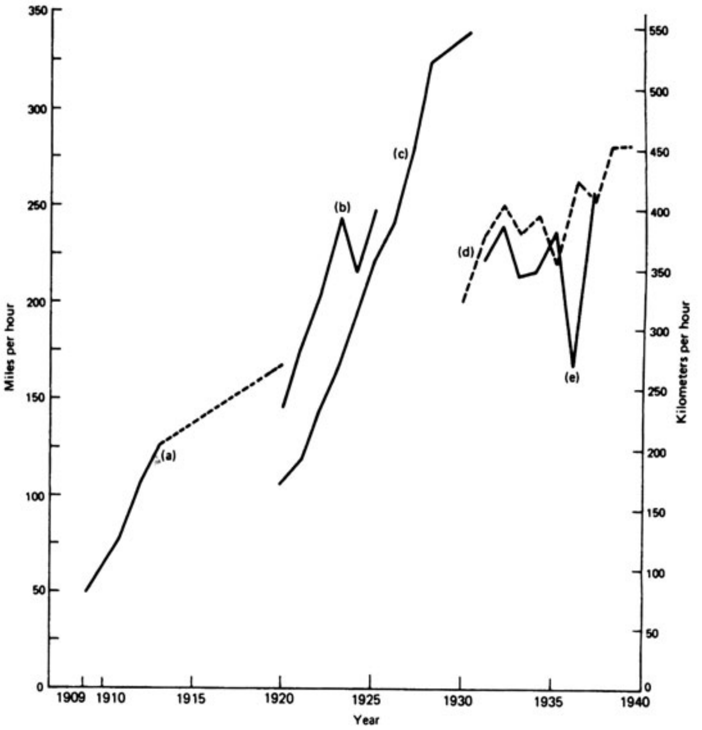


Fig. 15. Speed records of the major trophy races held during the period from 1909 to the 1930s. (a) The Gordon Bennett Cup; (b) The Schneider Cup; (c) The Pulitzer Trophy; (d) The Thompson Trophy; (e) The Bendix Trophy.

commissioned in mid-1942, but did not see action until mid-1944. Nearly 1,500 of these aircraft were produced, but only about 400 saw action during the war. The Sturmvogels were equipped with two jet engines, each capable of 1984 pounds (900 kilograms) of thrust. They had a speed of 541 miles (871 kilometers) per hour, a range of 650 miles (1046 kilometers), and a ceiling of a little over 37,500 feet (11,430 meters)—with a crew of one. Also of interest was the Messerschmitt Me-163B Komet. It was the first and, in fact, only rocket-propulsion fighter of the war. It was first commissioned at the end of July 1944. The craft had extremely limited performance parameters—flying time, about 10 minutes; altitude, about 6.2 miles (10 kilometers) (in 2.5 minutes). About 350 were produced and it is reported that more of these aircraft exploded upon landing than were knocked out by Allied fighters. The single power plant of the Komet was capable of producing about 3,750 pounds (1,701 kilograms) of thrust, with a maximum flying speed of just under 600 miles (965 kilometers) per hour. It carried a crew of one. The Japanese Yokosuka MXY7 Ohka 22, also known as the kamikaze or suicide aircraft, was equipped with a single jet engine in the tail. The little craft with a wing spread of under 14 feet (4.3 meters) and length of about 23 feet (7.0 meters), carried a 1,300-pound (590 kilograms) charge of dynamite. In operation, the Ohka was carried to near its target by a G4M2e, at which time the Ohka was piloted for a short distance to the target, with no escape for the pilot just prior to impact.

The war faced a challenge in refurbishing the fleets of not only the various airlines in the United States which had been utilized to the full during the war, but also of demands for aircraft by much of the rest of the world. Jet engine and jet technology, in general, still was not quite ready for application in large commercial planes and required further testing and refining in military aircraft, for which funding was available. Thus, during the late 1940s and early 1950s, a number of interesting and well-designed and performing propeller aircraft were introduced.

The first jet passenger service was offered by BOAC in early May 1952. The plane used was the DeHavilland DH 106 Comet. A couple of accidents in 1954 precipitated early retirement for these airplanes. However, in 1958, a revised version, the Comet 4, was introduced. About 75 of these airplanes were built. The early Comet 1 was powered by 4 jet engines, each with a thrust of 4,450 pounds (2019 kilograms). The wing spread was 115 feet (35 meters), the cruising speed was 490 miles (788 kilometers) per hour, with a ceiling of 35,000 feet (10,668 meters). The airplane could accommodate up to 44 passengers with a crew of 4.

The Boeing 707-121 made its first flight in mid-July 1954. However, it was not used for nonstop flights across the Atlantic until late 1958. This airplane generally is considered as one of the most successful passenger aircraft of all times. The early airplane was powered by 4 jet engines, each producing a thrust of 13,000 pounds (58 kN). The wing spread was 130 feet (39 meters), cruising speed 585 miles (941 kilometers) per hour, a range of 3,750 miles (6034 kilometers), and ceiling of 36,000 feet (10,973 meters). At maximum loading, it could accommodate 189 passengers with a crew of 4.

The Sud-Aviation SE-210 Caravelle made its first flight in late May 1955. It commenced commercial passenger service in Europe in 1959. Nearly 300 aircraft of the SE-210 were produced and, of course, there have been subsequent model changes. The SE-210 was powered by 2 jet engines, each with 10,500 pounds (47 kN) of thrust, wing spread was nearly 113 feet (34 meters), cruising speed 456 miles (734 kilometers) per hour, a range of 1150 miles (1850 kilometers), and ceiling of 30,000 feet (9144 meters). Maximum capacity was 80 passengers with a crew of 3.

In recent years, great emphasis has been placed upon the use of stronger, lighter-weight materials—both in military and passenger aircraft. See Figs. 16 and 17.

The problems of the supersonic passenger aircraft (*Concorde*) in terms of economic viability are well known. Periodically, interest is indicated in other supersonic passenger carriers that, because of the impressive advancements made in aircraft technology over the last several years, possibly could be economically and otherwise attractive to the commercial airlines.

Fig. 16. Airframe of the F/A-18 strike fighter aircraft is a balance of conventional materials and graphite epoxy providing an optimum lightweight structure. The wing skins, trailing edge flaps, stabilizors, vertical tails and rudder, speedbrake, and many access doors are made of graphite composite materials comprising about 9.9% of the aircraft structural weight and resulting in appreciable weight savings, plus increased aircraft performance. Other materials used (as a percent of structural weight) include aluminum, 49.6%; steel, 16.7%; titanium, 12.9%; other materials, 10.9%. (*McDonnell Douglas Corporation.*)

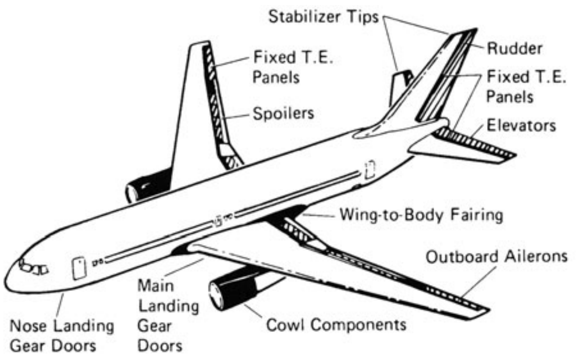


Fig. 17. Use of advanced composites on the Boeing 767. Solid black areas indicate hybrid composite (graphite-base); cross-hatched areas indicate graphite composite.

Additional Reading

Abelson, P. H.: "Jet-Powered Engines," *Science*, 497 (October 25, 1991).

Backman, D. G., and J. C. Williams: "Advanced Materials for Aircraft Engine Applications," *Science*, 1082 (February 28, 1992).

Beardsley, T.: "Earning Its Wings: Hypersonic Flight," *Sci. Amer.*, 17 (June 1988).

Goehler, D. D.: "Aerospace Industries Materials," *Advanced Materials and Processes*, 77 (January 1990).

Hull, D.: "Energy Absorbing Composite Structure," *University of Wales Review*, 23 (Spring 1988).

Jakab, P. L.: "Visions of a Flying Machine," Smithsonian Institution Press, Washington, D.C., 1990.

Jameson, A.: "Computational Aerodynamics for Aircraft Design," *Science*, 361 (July 28, 1989).

Kayton, M., Editor: "Navigation: Land, Sea, Air and Space," IEEE Press, New York, 1990.

Lukasak, D. A., and R. M. Hart: "Strong Aluminum Alloy Shaves Airframe Weight," *Advanced Materials and Processes*, 46 (October 1991).

March, A.: "The Future of the U.S. Aircraft Industry," *Technology Review (MIT)*, 26 (January 1990).

Nadis, S.: "The Answers Are Still Blowing in the Wind: Wind Tunnels," *Technology Review (MIT)*, 12 (January 1991).

Nelson, O. E.: "Titanium Staves Off Composites," *Advanced Materials and Processes*, 18 (June 1991).

Pallett, E. H. J.: "Automatic Flight Control," Blackwell Scientific Publications, Philadelphia, Pennsylvania, 1987.

Sims, C. T.: "Non metallic Materials for Gas Turbine Engines," *Advanced Materials and Processes*, 32 (June 1991).

Spitzer, "Digital Avionics Systems," Prentice-Hall, Hightstown, New Jersey, 1987.

Staff: "High-Flying Ideas," *Technology Review (MIT)*, 80 (October 1989).

Staff: "Advanced Materials in Aerospace Applications," ASM International, Materials Park, Ohio, 1990.

Staff: "Jane's All the World's Aircraft," Sampson Low, Merston & Co., Ltd., London (published annually).

Staff: "The Aerospace Year Book," Book Inc., Washington, D.C. (published annually).

Staff: "The National Plan for Aviation Human Factors," U.S. Department of Transportation, Federal Aviation Administration, Washington, D.C., November 1990.

Stephens, J. R.: "Composites Boost 21st-Century Aircraft Engines," *Advanced Materials and Processes*, 35 (April 1990).

Stix, G.: "Plane Geometry: Boeing Uses CAD," *Sci. Amer.*, 110 (March 1991).

Stix, G.: "Along for the Ride," *Sci. Amer.*, 94 (July 1991).

Stix, G.: "Smaller World. The Draper Prize Recognizes the Father of the Jet Age," *Sci. Amer.*, 57 (December 1991).

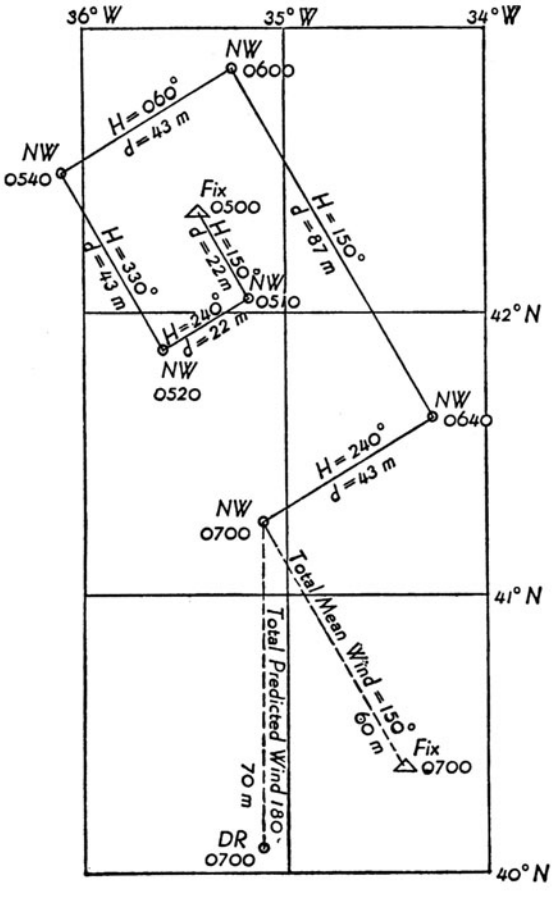
Suter, A. M.: "Noise Wars," *Technology Review (MIT)*, 47 (November 1989).

Wiener, T. L., and D. C. Nagel: "Human Factors in Aviation," Academic Press, San Diego, California, 1988.

AIR PLOT. A plot showing the movements of an airplane relative to the air is known, in navigation, as an air plot. It is similar in every respect to the old-fashioned method for obtaining the dead reckoning (DR) position of a sea-borne ship, disregarding the effects of ocean currents.

The air plot is made up of a series of distance vectors, added in the usual manner, the direction of each vector being a heading of the plane and the length being proportional to the distance moved through the air along the heading (distance = air speed × time on heading). The summation of the vectors gives a no wind (NW) position of the plane at any time. To obtain the DR position of the plane it is only necessary to add the total wind vector, whose direction is that of the motion of the air and whose length is the total movement (wind speed × total time since departure) of the air, to the NW position. Such a method avoids the necessity of constructing the velocity-vector diagram to obtain course and speed along course for each heading, as is necessary in the standard method for obtaining a DR position in air navigation.

In cases where the plane is frequently changing headings, the air-plot method saves from 30–60% of the time involved in obtaining the DR position by standard methods. Furthermore, in the frequent cases where the predicted wind for high altitude is very inaccurate, the air plot provides a method for determining the wind.



Scale figure for air plot.

The solution of the following problem indicates that advantages and value of the air plot. On a day when the predicted winds are uncertain, a plane is to search an area centered about a point in latitude L = 42°22′ N and longitude Lo = 35°25′ W. The plane, flying with true air speed TAS = 130 knots at altitude 10,000 feet (3,000 meters) is over the point at 0500 and heads 150°. The altitude and air speed are maintained, but the plane alters headings at the given times as follows: 0510 heading 240°, 0520 heading 330°, 0540 heading 060°, 0600 heading 150°, 0640 heading 240°. The air plot, properly labeled, is shown in the figure from which the NW position at 0700 is found to be L = 41°15′ N and Lo = 35°07′ W.

The predicted wind at the altitude of flight is from 000° and speed 35 knots. From the 0700 NW position a vector in direction 180° and of length 70 miles (112 kilometers) is drawn. This is shown as a dotted line in the figure and represents the total movement of the plane due to the predicted wind in the two hours from 0500. This gives a dead-reckoning position in latitude 40°05′ N and longitude 35°07′ W.

A fix is obtained at 0700 in latitude 40°23′ N and longitude 34°27′ W. If a vector is drawn from the NW position to the fix, it will represent the actual movement of the air in the interval from 0500 to 0700. From the diagram this is found to be in the direction 150° and 60 miles (96 kilometers) long. Since this represents 2 hours total motion of the air, we have a mean wind from 330°, speed 30 knots.

To obtain the DR position, using the standard methods of dead reckoning, would have required the plotting of at least four velocity-vector diagrams to obtain the course and ground speed on each heading. Each of these would have taken at least twice as much time as is required to plot an air vector and, after the courses and ground distances had been obtained, the geographic-motion vector diagram would have had to be plotted. Then to obtain the mean wind, the assumed air motion would have had to be subtracted from the DR position to obtain the NW position, which has to be used with the fix to obtain the actual total wind motion. See also **Course; Dead Reckoning; Heading;** and **Navigation.**

AIR POLLUTION. See **Pollution (Air).**

AIR PREHEATER. There are many devices of which the purpose is to heat air for some specific usage. However, in speaking of air preheaters, what is ordinarily meant is the heater employed for raising the temperature of air used for combustion of a fuel. This may occur in some industrial process such as preliminary heating of the air supplied to blast furnaces, but the most frequent use of air preheaters today is in connection with steam boilers. This type of air preheater is a heating surface installed between the boiler flue gas outlet and the stack. In arrangement the heating surface is composed either of tubes with flue gas inside and the air to be heated outside, or of rectangular plates spaced about one-half inch (13 millimeters) apart, leaving alternate gas and air passages. Its use is chiefly justified on economic grounds.

Two principles are employed for heat transfer in air preheaters. The recuperative principle implies transfer of heat through a separating partition, such as the walls of a tube, by continuously recuperating the cool side with conduction of heat from the hot side. Regenerative heaters are those which alternately heat and cool the same mass, regenerating it thermally by passing hot spent gas over its surface.

AIRSHIP. See **Dirigibles and Airships.**

AIRSPEED INDICATOR. Many air-data flight instruments operate from static or "pitot" pressure, or both. These pressures are usually obtained from a pitot-static tube of the type shown in Fig. 1. On the smooth side of the pitot tube, the pressure differs little from that of undisturbed air, and slots placed there provide a source of static pressure. Tubes usually are electrically heated to melt off any ice that might form and which would partly or fully seal off the openings and give erroneous readings in various air-data instruments. See Fig. 2. The difference between the pitot pressure and static pressure is a measure of *indicated airspeed*. Airspeed indicators are calibrated in knots (nautical mph) under standard pressure (29.92 inches of mercury) and temperature (59°F). A relationship between pressure difference and airspeed in knots is given in the accompanying table. See also **Pitot Tube.**

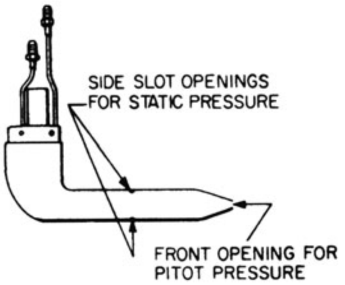


Fig. 1. Pitot tube of type used in aircraft.

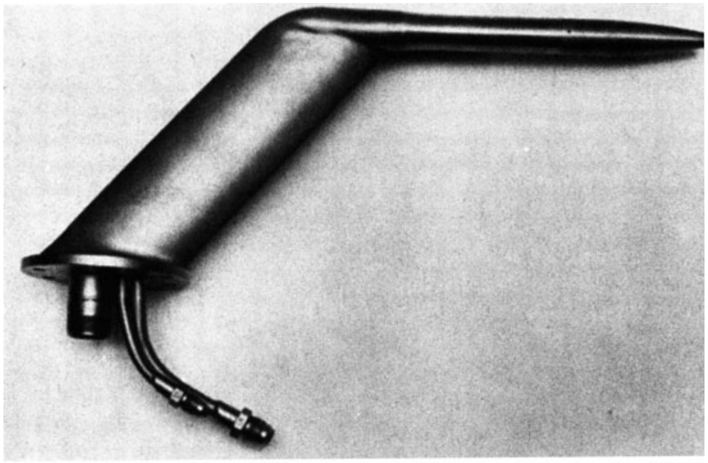


Fig. 2. Electrically heated pitot static tube.

AIRSPEED VERSUS DIFFERENTIAL PRESSURE AT STANDARD TEMPERATURE AND PRESSURE

Indicated Airspeed, Knots	Differential Pressure, in Hg	Indicated Airspeed, Knots	Differential Pressure, in Hg
0	0.000000	400	8.3973
10	0.004795	450	10.8837
20	0.019185	500	13.7967
30	0.043179	550	17.1859
40	0.076793	600	21.1088
50	0.1200	650	25.6316
60	0.1729	700	30.8159
70	0.2355	750	36.6276
80	0.3079	800	43.0092
90	0.3901	850	49.9241
100	0.4821	900	57.3481
125	0.7557	950	65.2644
150	1.0924	1,000	73.6613
175	1.4938	1,050	82.5302
200	1.9616	1,100	91.8649
225	2.4977	1,150	101.6606
250	3.1045	1,200	111.9136
275	3.7845	1,250	122.6210
300	4.5407	1,300	133.7805
350	6.2949	1,320	138.3706

SOURCE: Battelle Memorial Institute.

True-Airspeed Indicator. This instrument is primarily used for navigation. The measurement is a function not only of difference in pitot and static pressure, but also of temperature. In effect, a true-airspeed indicator (Fig. 3) is a computer operating on a rather complex relationship of two pressures and air temperature to give the correct indication. The relationship is:

$$V = 38.94\frac{M\sqrt{T_1}}{\sqrt{1 + 0.2KM^2}}$$

where V = true airspeed, knots; T_1 = indicated air temperature, Kelvin; K = recovery factor of temperature probe (usually 1.00 or slightly less); and M = Mach number.

Another approach is to use the speed of an air turbine as a measure of true airspeed, the intake pressure to the turbine being static pressure and the output being balanced to pitot pressure.

Machmeter. The ratio of true airspeed to the velocity of sound is called *Mach number*. The safe top speed of fast aircraft is usually expressed in terms of Mach number and is indicated on a Machmeter. See Fig. 4. This is also a mechanical computer which operates on the relationship of static and pitot pressures to give the proper indication. The relationship for Mach number is:

$$\frac{dP}{Ps} = (1 + 0.2M)^{23.5} - 1 \quad \text{for } M < 1$$

$$\frac{dP}{Ps} = \frac{166.9215M^7}{(7M^2 - 1)^{2.5}} - 1 \quad \text{for } M > 1$$

where M = Mach number; dP = differential (pitot less static) pressure; and Ps = static pressure. The true airspeed and Mach number are combined in Fig. 4.

Fig. 4. Combined true-airspeed and Mach number computer indicator.

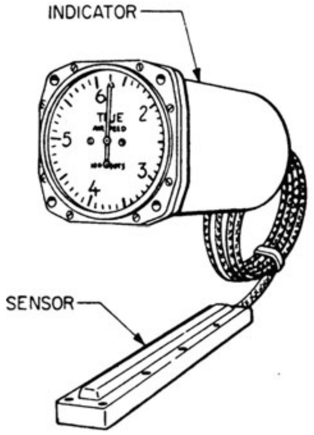


Fig. 3. True-airspeed indicator.

ALABANDITE. Manganese sulfide, MnS. Associated with pyrite, sphalerite, and galena in metallic sulfide vein deposits.

ALABASTER. A fine-grained variety of the mineral gypsum, formerly much used for vases and statuary. It is usually white in color or may be of other light, pleasing tints.

The word alabaster is derived from the Greek name for this substance. See also **Gypsum.**

ALBEDO (Astronomy). A term used to indicate the reflecting power of an object. Technically defined, albedo (often referred to as the Bond albedo) is the ratio of the radiation reflected from an object to the total amount incident upon it. Thus,

$$A = pq$$

where p is the ratio of the brightness at the phase angle of zero to the brightness of a perfectly diffusing disk under the same conditions, and where q is a factor representing the phase law. For example, the albedo of the moon is 0.068, which means that the moon reflects that fraction of the sunlight which is incident upon it. The value of the albedo of a planet is a measure of the quantity of atmosphere that surrounds the object. The higher the albedo, the thicker the atmospheric layer. In the case of objects without atmosphere, as in the case of the moon, the albedo, combined with the color of the reflected light, may be used to make estimates of the character of the material making up the surface of the object.

ALBEDO (Physics). In nuclear physics, the albedo is the ratio of the neutron current density out of a (non-source) medium to the neutron current density into it. In cosmic ray physics, albedo refers to those energetic, charged secondary cosmic rays that move generally upward. Because some of these particles go far into the upper atmosphere and return to the earth's surface (due to the effect of the earth's magnetic field), measurements of primary cosmic radiation must be corrected for the contribution of the albedo particles.

ALBERTITE. An oxygenated hydrocarbon which differs from asphaltum slightly in that it is not completely soluble in turpentine, nor can it be perfectly fused. Specific gravity, 1.097, pitchy luster, dark brown to black color. Occurs in veins from 1 to 16 feet (0.3 to 5 meters) wide in the Albert Shale of Albert County, New Brunswick.

ALBINISM. Absence of pigmentation. The condition has been noted in occasional individuals of many species which are normally pigmented, including humans. In some cases the term is applied to a partial lack of pigment, as in the white form of certain normally yellow butterflies; in this form the black markings characteristic of the species are fully developed. In contrast, the total lack of pigment in albino birds and mammals is shown by the pink eyes. In these organs pigment is functionally important but the albino fails to develop it, hence the color of the blood is seen through the tissues. Albinism in humans is known to be inherited as a recessive to normal pigmentation. The term is also applied to normally green plants which fail to develop chlorophyll. As in humans, albinism is inherited as a simple recessive. Albino plants die as soon as food is exhausted from the seed. Partial albinism in plants is known as variegation.

Skin pigmentation and tanning are related to the amount of melanin in epidermal melanosomes. In the latter, tyrosine is oxidized to dopa and dopaquinone by the enzyme *tyrosinase*. The intermediate compounds polymerize to form the brown-black melanin pigment. Malfunction of this process, rather than resulting in what is considered normal pigmentation for a given species, may result in hypo- or hyperpigmentation.

Albinism results from a deficiency in melanin formation and is a form of hypopigmentation. *Hypomelanosis* may result from trauma or certain infections, including leprosy, syphilis, and pinta. *Oculocutaneous albinism*, involving white skin and hair, photophobia, and nystagmus, is noted at birth. This is the result of an inborn error of metabolism, inherited as an autosomal recessive trait. The occurrence is rare, estimated at one case in 20,000 individuals. In the most common form of albinism, the individual cannot synthesize tyrosinase. Individuals with photophobia should use sunglasses and essentially avoid exposure to the sun. Medications incorporating sunscreening substances can be used.

A condition known as *vitiligo* is encountered in about one percent of individuals. It has been suggested that vitiligo is probably inherited as an autosomal dominant characteristic. Areas of macular depigmentation occur on the extremities, face, neck, and skin folds. Bilateral symmetry is common. Although vitiligo usually is found in individuals who otherwise have normal health, some patients with vitiligo are prone to such disorders as thyroid disease (hyperthyroidism, Grave's disease, thyroiditis), Addison's disease, pernicious anemia, and alopecia areata. There may be an association of vitiligo with diabetes. A number of relatively useful remedial treatments are available to alleviate the cosmetic defects of vitiligo, including avoidance of exposure to the sun, use of sunscreens, the topical or oral use of methoxsalen (photosensitizing drug), triosalen, followed by controlled exposure to sunlight or long-wave ultraviolet light.

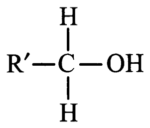
ALBUMIN. An albumin is a member of a class of proteins which is widely distributed in animal and vegetable tissues. Albumins are soluble in water and in dilute salt solutions, and are coagulable by heat.

Albumin is of great importance in animal physiology; in man it constitutes about 50% of the plasma proteins (blood) and is responsible to a great extent for the maintenance of osmotic equilibrium in the blood. The high molecular weight (68,000) of the albumin molecule prevents its excretion in the urine; the appearance of albumin may indicate kidney damage. See **Kidney and Urinary Tract.**

ALCOHOL. A term commonly used to designate ethyl alcohol or ethanol. See **Ethyl Alcohol.** Also a class of organic compounds. See **Alcohols.**

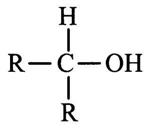
ALCOHOLATE. Replacement of the hydrogen in the hydroxyl group of an alcohol by a metal, particularly a metal that forms a strong base, results in formation of an alcoholate. An example is sodium ethylate, C_2H_5ONa.

ALCOHOLS. The alcohols may be regarded as hydrocarbon derivatives in which the hydroxyl group (OH) replaces hydrogen on a saturated carbon. Alcohols are classified as *primary*, *secondary*, or *tertiary*, according to the number of hydrogen atoms that are bonded to the carbon atom with the hydroxyl substituent. Alcohols also may be regarded as alkyl derivatives of water. Thus, alcohols with a small hydrocarbon group tend to be more like water in properties than a hydrocarbon of the same number. Alcohols with a large hydrocarbon group are found to have physical properties similar to a hydrocarbon of the same structure. Some comparisons are given in the accompanying table. Structures are summarized by:



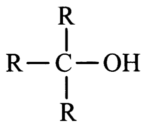
$$R'-\overset{\displaystyle H}{\underset{\displaystyle H}{C}}-OH$$

Primary



$$R-\overset{\displaystyle H}{\underset{\displaystyle R}{C}}-OH$$

Secondary



$$R-\overset{\displaystyle R}{\underset{\displaystyle R}{C}}-OH$$

Tertiary

where $R' =$ H, alkyl, aryl; $R =$ alkyl, aryl.

In addition to the basic classification as primary, secondary, or tertiary, alcohols may be further grouped according to other structural features. *Aromatic alcohols* contain an aryl group attached to the carbon having the hydroxyl function; *aliphatic alcohols* contain only aliphatic groups. The prefix *iso* usually indicates branching of the carbon chain.

Alcohols containing two hydroxyl groups are called *dihydric alcohols* or *glycols*. Ethylene glycol, $HOCH_2CH_2OH$, trimethylene glycol, $HOCH_2CH_2CH_2CH_2OH$, and 1,4-butanediol are examples of industrially important glycols. Glycerol, $HOCH_2CHOHCH_2OH$, has three hydroxyl groups per molecule and is a *trihydric* alcohol. Physical properties of alcohols containing more than one hydroxyl group can be estimated by considering the number of carbons for each hydroxyl group as in the case of simple alcohols.

Reactions of Alcohols. Alcohols undergo a large number of reactions. However, these reactions may be grouped into a few general types. Reactions of alcohols may involve the O—H or C—O bonds.

TABLE 1. COMPARISON OF PHYSICAL PROPERTIES OF ALCOHOLS AND HYDROCARBONS

Alcohol	Hydrocarbon	Formula	Properties
Methanol		CH_3OH	Liquid, water soluble, bp. 65° C
	Methane	$CH_3—H$	Gas, water insoluble
Ethanol		CH_3CH_3OH	Liquid, water soluble, bp. 78.5° C
	Ethane	$CH_3CH_2—H$	Gas, water insoluble
Tetradecanol		$CH_3(CH_3)_{12}CH_2OH$	Liquid, water insoluble, bp. 263.2° C
	Tetradecane	$CH_3(CH_2)_{12}CH_2—H$	Liquid, water insoluble, bp. 253.5° C

Ester formation and salt formation are examples of the former class, while conversion to halides is an example of the latter type.

—O—H Bond Cleavage

(1) $ROH + CH_3COOH \overset{H_+}{\rightleftharpoons} CH_3COOR + H_2O$
(2) $ROH + K \rightarrow RO^-K^+ + \frac{1}{2}H_2$

C—O Bond Cleavage

(3) $RCH_2OH + HCl \xrightarrow[170^\circ C]{ZnCl2} RCH_2Cl + H_2O$

Many industrially important substitution reactions of alcohols are conducted in the vapor phase over a catalyst. Only primary alcohols give satisfactory yields of product under these conditions.

(4) $CH_3OH + H_2S \xrightarrow{K_2WO_4} CH_3SH + H_2O$
(5) $RCH_2OH + (CH_3)_2NH \xrightarrow{Al_2O_3} RCH_2N(CH_3)_2 + H_2O$

Production of Alcohols. Lower alcohols (amyl and below) are prepared by (a) hydrogeneration of carbon monoxide (yields methanol), (b) olefin hydration (yields ethanol, isopropanol, secondary and tertiary butanol), (c) hydrolysis of alkyl chlorides, (d) direct oxidation, and (e) the OXO process.

(6) $C{=}O + 2H_2 \rightarrow CH_3OH$
(7) $CH_3CH{=}CH_2 + H_2O \overset{H^+}{\rightarrow} CH_3CHOHCH_3$
(8) $C_5H_{11}Cl + H_2O \rightarrow \underset{\text{(mixture of isomers)}}{C_5H_{11}OH}$

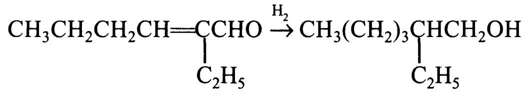
Most higher alcohols (hexanol and higher) and primary alcohols of three carbons or more are synthesized by one of four general processes, or derived from a structurally related natural product. See also **Organic Chemistry.**

The OXO Process. An olefin may be hydroformylated to a mixture of aldehydes. The aldehydes are readily converted to alcohols by hydrogenation. Many olefins from ethylene to dodecenes are used in the OXO reaction. OXO alcohols are typically a mixture of linear and methyl branched primary alcohols.

(9) $RCH{=}CH_2 + CO + H_2$
$\rightarrow [RCH_2CH_2CHO + RCH(CHO)CH_3]$

Aldol Condensation. Aldehydes may also be dimerized by an aldol condensation reaction to give a branched unsaturated aldehyde. This may be converted to a branched alcohol by hydrogenation.

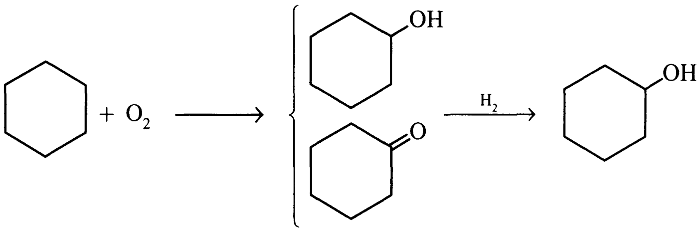
(10)



$$2CH_3CH_2CH_2CHO \rightarrow CH_3CH_2CH_2CH{=}\underset{\displaystyle C_2H_5}{\underset{|}{C}}CHO \overset{H_2}{\rightarrow} CH_3(CH_2)_3\underset{\displaystyle C_2H_5}{\underset{|}{C}}HCH_2OH$$

Alcohols from an aldol reaction may be linear if acetaldehyde is a reactant, but usually aldol alcohols are branched primary alcohols. An aldol condensation sometimes is done with an OXO reaction. The combined process is called the ALDOX process.

Oxidation of Hydrocarbons. Using air, the oxidation of hydrocarbons generally results in a mixture of oxygenated compounds and is not a useful synthesis of alcohols except under special circumstances. Cyclohexanol may be prepared by air oxidation of cyclohexane inasmuch as only one isomer can result.



The yield of alcohol from normal paraffin oxidation may be improved to a commercially useful level by oxidizing in the presence of boric acid.

(12) $3RH + \frac{3}{2}O_2 + H_3BO_3 \rightarrow (RO)_3B + 3H_2O$
(13) $(RO)_3B + 3H_2O \rightarrow H_3BO_3 + 3ROH$

A borate ester is formed which is more stable to further oxidation than the free alcohol. This is easily hydrolyzed to recover the alcohol. These alcohols which are predominately secondary are used in surfactant manufacture.

Synthesis from Alkylaluminums. Fundamental work on organoaluminum chemistry by Prof. Karl Ziegler and co-workers at the Max Planck Institute provided the basis for a commercial synthesis of even-carbon-numbered straight chain primary alcohols. These alcohols are identical with products derived from naturally occurring fats. In this process, ethylene is reacted with aluminum triethyl to form a higher alkylaluminum which then is oxidized and hydrolyzed to give the corresponding alcohols.

(14) $(C_2H_5)_3Al \xrightarrow{CH^2{=}CH^2}$
$[CH_3(CH_2CH_2)_nCH_2]_3Al \xrightarrow[(2)\,H_2O]{(1)\,O_2}$
$3CH_3(CH_2CH_2)_nCH_2OH + Al(OH)_3$

Commercialization of this route to higher alcohols is the most significant development in this area in recent years.

Synthesis from Natural Products. Many alcohols are prepared by reduction of the corresponding methyl esters which are derived from animal or vegetable fats. These alcohols are straight chain even-carbon-numbered compounds. Tallow and coconut oil are two major raw materials for higher alcohol manufacture.

(15) $(RCOO)_3C_3H_5 + 3CH_3OH \rightarrow 3RCOOCH_3$
Triglyceride
$+ CH_2OHCHOHCH_2OH$
(16) $RCOOCH_3 + 2H_2 \xrightarrow{Catalyst} RCH_2OH + CH_3OH$

The production of ethyl alcohol for beverage, cosmetic, and pharmaceutical products is commonly accomplished by the natural process of fermentation. See **Ethyl Alcohol;** and **Fermentation.** Beer, wine, and whiskey production are extensively covered in the *Foods and Food Production Encyclopedia* (D. M. and G. D. Considine, Eds.), Van Nostrand Reinhold, New York, 1982. Methyl alcohol is described in detail in entry on **Methanol.**

ALCOHOLYSIS. If a triglyceride oil is heated with a polyol, such as glycerol or pentaerythritol, mixed partial esters are produced in a reaction known as *alcoholysis*.

ALCYONARIA. An order of the class *Anthozoa* (*Actinozoa*) in the phylum *Coelenterata*. The animals of this order are marine and because of the hard deposits formed in and around their bodies they are often known as corals. Among them are the sea-pen, the sea-fan, organ-pipe coral, and precious coral.

Alcyonaria are usually found in colonies. The individuals are polyps connected together by living structures and by the hard skeletal structures. They differ from the true corals in having only eight tentacles, pinnately branched. See also **Coelenterata.**

ALDEBARAN (α Tauri). A star whose name is derived from an Arabic phrase indicating that the star is the "leader of the followers," i.e., the leader of the asterism known as the Hyades, which follow the Pleiades in their nightly journey across the sky. Astrologically, Aldebaran was a fortunate star, portending riches and honor. This star was one of the four royal stars of the Persians about 3000 B.C.

Aldebaran is one of the smaller stars, whose diameter has been measured with the stellar interferometer. The diameter is found to be about 6.4×10^7 kilometers, or 46 times the diameter of our sun. Ranking fourteenth in apparent brightness among the stars. Aldebaran has a true brightness value of 100 as compared with unity for the sun. Estimated distance from the earth is 53 light years. Aldebaran is classified as a star of orange color and of spectral type K. See also **Constellations;** and **Star.**

ALDEHYDES. The homologous series of aldehydes (like ketones) has the formula $C_nH_{2n}O$. The removal of two hydrogen atoms from an alcohol yields an aldehyde. Thus, two hydrogens taken away from ethyl alcohol $CH_3{\cdot}C(H_2)OH$ yields acetaldehyde CH_3CH_2OH; and two hydrogens removed from propyl alcohol $C_2H_5{\cdot}C(H_2)OH$ yields propaldehyde $C_2H_5{\cdot}CHO$. The trivial names of aldehydes derive from the fatty acid which an aldehyde will yield upon oxidation. Thus, formaldehyde is named from formic acid, the latter being the oxidation product of formaldehyde. Similarly, acetaldehyde is oxidized to acetic acid. Or, the aldehyde may be named after the alcohol from which it may be derived. Thus, formaldehyde which may be derived from methyl alcohol may be named methaldehyde; or acetaldehyde may be named ethaldehyde since it may be derived from ethyl alcohol. In still another system, the aldehyde may take its name from the parent hydrocarbon from which it theoretically may be derived. Thus, propa*nal* (not to be confused with propa*nol*) may signify propaldehyde (as a derivative of propane).

Essentially aldehydes exhibit the following properties: (1) with exception of the gaseous formaldehyde, all aldehydes up to C_{11} are neutral, mobile, volatile liquids. Aldehydes above C_{11} are solids under usual ambient conditions; (2) formaldehyde and the liquid aldehydes have an unpleasant, pungent, irritating odor, (3) although the low-carbon aldehydes are soluble in H_2O, the solubility decreases with formula weight, and (4) the high-carbon aldehydes are essentially insoluble in H_2O, but are soluble in alcohol or ether.

The presence of the double bond (carbonyl group C:O) markedly determines the chemical behavior of the aldehydes. The hydrogen atom connected directly to the carbonyl group is not easily displaced. The chemical properties of the aldehydes may be summarized by: (1) they react with alcohols, with elimination of H_2O, to form *acetals*; (2) they combine readily with HCN to form *cyanohydrins*, (3) they react with hydroxylamine to yield *aldoximes*; (4) they react with hydrazine to form *hydrazones*; (5) they can be oxidized into *fatty acids* which contain the same number of carbons as in the initial aldehyde; (5) they can be reduced readily to form *primary alcohols*. When benzaldehyde is reduced with sodium amalgam and H_2O, benzyl alcohol $C_6H_5{\cdot}CH_2{\cdot}OH$ is obtained. The latter compound also may be obtained by treating benzaldehyde with a solution of cold KOH in which benzyl alcohol and potassium benzoate are produced. The latter reaction is known as Cannizzaro's reaction.

In the industrial production of higher alcohols (above butyls), aldehydes play the role of an intermediate in a complete process that involves aldol condensation and hydrogenation. In the OXO process, olefins are catalytically converted into aldehydes that contain one more carbon than the olefin in the feedstock. Aldehydes also serve as starting materials in the synthesis of several amino acids. See also **Acetaldehyde; Aldol Condensation; Benzaldehyde;** and **Furfuraldehyde.**

ALDERFLY (*Insecta, Neuroptera*). The alderflies belong to the order *Neuroptera*, suborder *Megaloptera*. These are the most primitive members of the order, and the most primitive of all holometabolous insects. There are only two members of the *Megaloptera* in Europe, both of which are Alderflies (family *Sialidae*). These are small, with a wingspan of only 3 centimeters (1 inch). *Sialis lutaria* is commonly found in central Europe in the spring or summer, sitting on shore plants by still or slowly flowing waters. The second species, *Sialis fuliginosa*, is usually rather more dusky in color and differs from the first in the wing venation and the structure of the hind end of the body. The Smoky Alderfly (*Sialis infumata*) of the U.S., with brownish-black wings, is similar in size and habits to the European species. At rest, the alderflies lay their wings together over the body like a roof, as do almost all the *Neuroptera*. They are hunters, both as adults and as larvae. The larvae, however, are aquatic.

ALDER TREES. Members of the family *Betulaceae* (birch family), these trees are of the genus *Alnus*. They are deciduous trees or shrubs and are considered very hardy.

The red alder is found in forests on the west coast of the United States and Canada, from California northward to Alaska. The average height is from 40 to 60 feet (12 to 18 meters). However, as will be noted from the accompanying table of record alder trees, the red alder can reach heights of close to 100 feet (30 meters). The white alder occurs on the Pacific slopes and can attain a height of over 100 feet (30 meters), although the average is from 50 to 70 feet (15 to 21 meters). Staminate flowers are contained in catkins of about 5 inches (13 centimeters) in length. The cones range from $\frac{1}{2}$ to $\frac{7}{8}$ inch (1.2 to 2.2 centimeters) in length. The leaves are ovate and of a light yellow-green color. The tree ranges from northern Idaho westward to southeastern Oregon, with preference for the eastern slopes of the Cascade Mountains. The tree also is found in the Coastal Range and Sierra Nevadas down into southern California.

Alder wood is durable and well suited for applications in wet and damp areas, as may be required for sashes and bridge construction. In areas where forest fires have cleared the land, alders are fast-growing and come up abundantly. Alders have nodules which form on the root and which help to enrich the soil. However, if hemlock and Douglas fir are present, these latter trees can soon overtake the alder in height, smothering out all except the very strongest trees. Alder lumber is usually shipped from west coast forests. In the green condition, red alder has a moisture content of 98% and weighs 46 pounds per cubic foot (737 kilograms per cubic meter). When air-dried to 12% moisture content, the weight is 28 pounds per cubic foot (448.5 kilograms per cubic meter), with 1,000 board-feet (2.36 cubic meters) of nominal sizes weighing about 2330 pounds (1057 kilograms). Crushing strength when compression is applied parallel to the grain is 2960 psi (20.4 MPa) for the green wood; 5820 psi (40.1 MPa) for the dry wood. Tensile strength when tension is applied perpendicular to the green wood is 390 psi (2.7 MPa) for the green wood, 420 psi (2.9 MPa) for the air-dried wood.

Record alder trees in the United States are listed in the accompanying table.

RECORD ALDER TREES IN THE UNITED STATES[1]

Specimen	Circumference[2]		Height		Spread		Location
	Inches	Centimeters	Feet	Meters	Feet	Meters	
Arizona alder (1982) (*Alnus oblongifolia*)	199	505	129	39.3	50	15.2	New Mexico
European alder (1982) (*A. glutinosa*)	131	333	70	21.3	40	12.2	Illinois
Green alder (1984) (*A. crispa*)	10	25	28	8.5	12	3.7	Michigan
Mountain alder (1984) (*A. tenuifolia*)	86	218	50	15.2	31	9.4	Washington
Red alder (1980) (*A. rubra*)	245	622	104	31.7	49	14.9	Oregon
Seaside alder (1975) (*A. maritima*)	13	33	28	8.5	11	3.4	Virginia
Sitka alder (1967) (*A. sinuata*)	21	53	30	9.1	—	—	Oregon
Smooth alder (1985) (*A. serrulata*)	14	36	37	11.3	24	7.3	Texas
Speckled alder (1984) (*A. rugosa*)	38	97	66	20.1	56	17.1	Michigan

[1] From the "National Register of Big Trees," The American Forestry Association (by permission).
[2] At 4.5 feet (1.4 meters) from ground level.

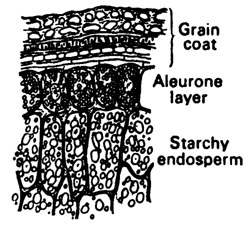
ALDOL CONDENSATION. A reaction between aldehydes or aldehydes and ketones which occurs without the elimination of any secondary product, and yields β-hydroxycarbonyl compounds. It is distinguished from polymerization by the fact that it occurs between aldehydes and ketones and is not generally reversible. In its simplest form it may be represented by the condensation of two molecules of acetaldehyde to aldol:

$$CH_3{\cdot}CHO + CH_3{\cdot}CHO \rightarrow CH_3{\cdot}CHOHCH_2{\cdot}CHO$$

Weak alkalies and acids are employed to effect the condensation.

Researchers at the University of California, Berkeley, have accomplished acyclic stereocontrol through the aldol condensation. As observed by C. H. Heathcock (*Science*, **214**, 295–400, Oct. 23, 1981), one of the most difficult problems in the synthesis of complex organic compounds is that of controlling the relative stereochemistry, that is, establishing the correct configuration at the various chiral centers as the synthesis is carried out. In recent years, researchers have been attempting to find direct solutions to the problem, particularly in synthesizing acyclic and other conformationally flexible molecules. Heathcock and colleagues have found that aldol condensation, one of the oldest and most famililar organic reactions, can be a very effective tool for achieving stereocontrol.

ALEURONE GRAINS. Protein reserves found in the seeds of several different kinds of plants. In many plants there is a special aleurone layer of definite thickness found in the endosperm. In corn (maize), the layer is a single cell in thickness and may contain a colored pigment; in oats, it is two cells thick. In the castor oil plant, the aleurone grains are not restricted to a single layer, but are distributed rather generally in the endosperm, and have a complex structure. See accompanying figure.



The aleurone layer in wheat. Part of a section of a wheat grain showing aleurone layer with cells filled with granules of protein.

ALEWIFE (*Osteichthyes*). A common fish (*Pomolobus pseudoharengus*) of the Atlantic coast. It is related to the herring and shad. It enters the streams to spawn and is found in lakes of New York. The alewife is generally considered an unsatisfactory food fish. Normally found in marine waters, Nova Scotia to Florida. Maximum length about 11 inches (28 centimeters).

ALEXANDRITE. A variety of chrysoberyl, originally found in the schists of the Ural Mountains. It absorbs yellow and blue light rays to such an extent that it appears emerald green by daylight but columbine-red by artificial light. It is used as a gem, and was named in honor of Czar Alexander II of Russia. See also **Chrysoberyl.**

ALFALFA. Of the family *Leguminosae* (pea family), alfalfa (*Medicago sativa*), is a plant native to southwestern Asia. It is an important perennial forage plant, extensively cultivated since the Roman civilization. The plant has a very deeply penetrating root system, reaching down 25 feet (7.5 meters) or more and thus is well suited for growing in dry lands where resistance to drought is important. The fragrant, purplish flowers are a good source of honey. The plants are cultivated principally for hay. See also **Leguminosae.**

ALFVÉN WAVE. A transverse wave in a hydromagnetic field in which the driving force is the tension introduced by the magnetic field along the lines of force. The dynamics of such waves are analogous to those in a vibrating string, the phase speed being given by

$$C^2 = \mu H^2/4\pi\rho \tag{1}$$

where μ is the permeability, H the magnitude of the magnetic field, and ρ the fluid density. Dissipative effects due to fluid viscosity and electrical resistance may also be present.

For proton-electron plasmas, a simpler expression than that of Equation (1) may be obtained in terms of the magnetic field strength, B, and the number of charged particles, n, per unit volume. This expression gives the phase velocity as

$$C = \frac{(1.7 \times 10^6)B}{n^{1/2}}$$

The Alfvén wave behaves like a conventional shock wave, such as a supersonic sound wave, in having a high energy front resulting from an excess of the plasma velocity over the Alfvén velocity. However, it differs in that it is produced by the interaction of waves instead of particles, since the latter, in, for example, a plasma in the upper atmosphere or in space, are too widely spaced to have an appreciable number of collisions.

ALGAE. Thallophytes characterized by possessing chlorophyll, and so capable of elaborating their food by photosynthesis. Often the green pigment is completely concealed by other pigments, so that the plant is brown, red, or even black.

Algae are found in almost every habitat. In the oceans vast numbers of minute species float suspended in the upper levels of the water, while the shores are covered with many and varied forms from high tide level to depths of 30 feet (9 meters) or more. In fresh water they are equally abundant, but due to their smaller size are seldom so conspicuous as the marine forms; they occur in running water, in ponds, and in stagnant, often putrid, water. Many species are found only in hot springs. They are found on the surface of the ground, on the bark of trees, on rocks, and even underground to a depth of several feet. A few species have found a favorable habitat within the bodies of higher plants and animals. In fact, wherever they find support and can obtain the necessary materials for growth, there algae may be found.

Algal plants offer a wonderful diversity of forms. In size they range from unicellular microscopic plants to structures having dimensions comparable to the larger land plants. The plankton forms, those free-floating plants often so abundant in both fresh and salt water, are nearly all unicellular; other free-floating forms, usually found near shore or in small fresh water ponds and streams, are multicellular organisms of various shapes, filamentous forms being especially common. Finally, attached marine forms often attain massive dimensions. The common kelp, or devil's apron, of the colder coastal waters of North America may grow to a length of 30 feet (9 meters) or more, and to a width of 2 or 3 feet (0.6 or 0.9 meters), while related species found in the Pacific Ocean far exceed them in size, reaching lengths of 100 feet (30 meters) and more.

Not only do algae vary greatly in size, but they also show almost every conceivable shape. Unicellular types are often adorned with a complex but beautifully symmetrical series of arms, or bristles, which may be of service in keeping them floating in the water. The filamentous forms may be simple or very much branched; often they are delicate plants of rare beauty. Other algae grow in flat sheets or membranes, either spreading over the substratum or rising gracefully in the water. The larger forms are of coarser habit, varying from irregular tumorous plants to long slender cords and broad flat fronds. Plants of the genus *Sargassum*, one of the brown algae found in warm regions, have an appearance very similar to that of flowering plants. Each plant has a slender branching stalk, often 2 feet (0.6 meters) or more in length. From this stalk flat lateral branches arise, which look very much like leaves, except for their brown color. Other short lateral branches end in small sub-spherical balls easily mistaken for fruits. Actually these structures are hollow bladders which help to keep the plant floating. Other branches, the real reproductive parts of the plant, are short cylindrical objects which might be mistaken for buds. At first the plant grows on rocks and other solid objects. It is easily broken loose, however, and floats about in the ocean currents. Thus, these plants are frequently washed up on northern beaches.

There are several systems of classification of algae, varying in details but all using the various pigments found in their cells as a basis for separation. Obviously such a classification is very artificial, but in the algae it seems to agree quite closely with natural systems based on such other criteria as the structure of the thallus or plant body, as the substances formed by the cells and stored in them, and especially as the reproductive processes which are found in the different groups.

The algae are now commonly separated into seven divisions or phyla, although in an older system these groups had the status of Class. The four larger phyla are the *Cyanophyta* or blue-green algae. *Chlorophyta* or green algae, *Phaeophyta* or brown algae, and the *Rhodophyta* or red algae. The smaller phyla are the Euglenophyta or Flagellates, the Chrysophyta, including the yellow-green and golden-brown algae and the diatoms, and the Pyrrophyta or cryptomonads.

The blue-green algae are characterized by having within the cell, in addition to chlorophyll, a bluish pigment, phycocyanin. These pigments are not localized in a definite pigment-bearing body or plastid, but are frequently diffused in the outer zone of protoplasm. Surrounding the protoplasm is a definite cell wall of cellulose (see **Carbohydrates**). In most blue-green algae the outer part of the wall is modified and becomes a soft slimy substance which often forms a layer of considerable thickness. This slime substance may be of great value as an insulation against heat and desiccation, thus enabling blue-green algae to live in what seem to be most unfavorable environments. In the central portion of the protoplast are found many chromatin bodies, which, however, are not organized into a definite nucleus. The structure of the cell of these algae, with its absence of plastids and any definite nucleus, seems to indicate a relatively primitive organism, and is suggestive of the structure found in bacteria. See Fig. 1.

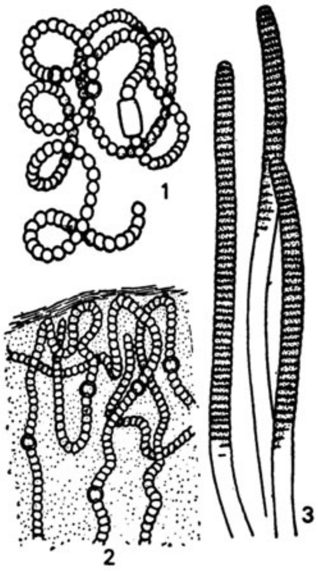


Fig. 1. Three species of *Myxophyceae*: (1) filaments of a species of *Anabaena*; (2) filaments of *nostoc* encased in the characteristic jelly; (3) *Oscillatoria*. (*After Smith and Fremy*)

Many blue-green algae are single-celled organisms. The individual cells are often separate or they may be held together by the gelatinous outer wall in aggregates sometimes of considerable size. Representatives of this group are frequently observed in temporary puddles formed by a summer shower or in quiet shallow ponds in which the water often becomes very warm. Some of them are of considerable economic importance, when they appear in water supply reservoirs. In these they sometimes occur in numbers so great as to color the water and to be only too obvious to the most casual observer. Such occurrences are frequently described as "water-blooms." Due to the products formed by the metabolism of the cells and liberated into the water, such "blooms" are real problems, for not only do these substances give to the water a distinctly unpleasant oily fishy taste and color, but several cases are recorded in which drinking of such water has been quickly fatal to livestock.

Other blue-green algae occur as simple filaments of cells. Single filaments may occur among other algae or they may exist in extensive masses covering considerable areas with a soft felt-like layer. In some genera many filaments are held together in a common gelatinous sheath. Many blue-green genera show what is called false branching. The rapid division of cells in the middle of a filament causes them to grow out laterally. Sometimes the filaments grow out as a single branch, as in *Tolypothrix*, or in pairs, as in *Scytonema*. True branching is found in a few genera.

In all the non-filamentous species the only method of reproduction is that of cell division. In the filamentous forms continued division may produce a filament of indefinite length. However, it eventually breaks up. This fragmentation may be due to animals feeding on cells of the filament, or to the death of certain cells. In some forms it is due to the development of cells which do not adhere tenaciously to the cells adjoining them. Specialized cells known as heterocysts are largely responsible for the last condition. They are large transparent cells which develop from ordinary cells. See Fig. 2.

Blue-green algae, particularly abundant in regions having a warm climate, develop elsewhere in great abundance, during the warmer sea-

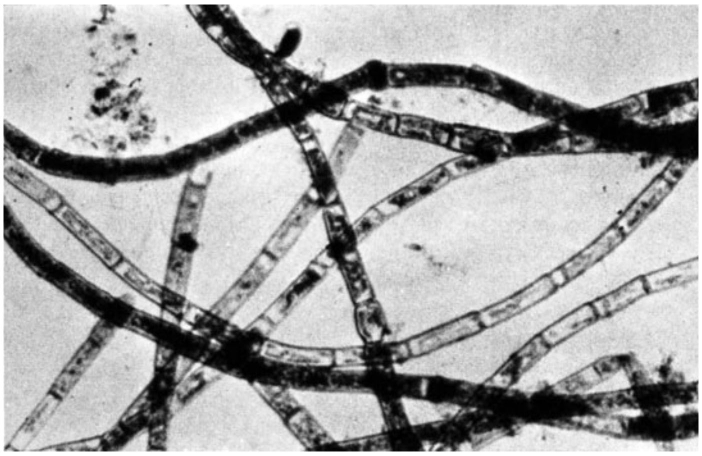


Fig. 2. *Oedogonium*. Filamentous cells in fresh water. (*Photomicrograph by B. J. Ford; copyright*)

sons of the year. Besides being a source of trouble in water reservoirs, blue-green algae sometimes form unsightly stains by growing on the walls of stone buildings, especially if the latter be constantly wet. Some of the blue-green algae have the ability to fix nitrogen (see **Bacteria**). It has been suggested that this may be important to such crops as rice.

The green algae, or Chlorophyta, are found in both salt and fresh water, where they often form conspicuous masses. They are characterized by a bright green color. The reserve food stored by these algae is starch, which is usually found around certain bodies known as pyrenoids, located in the chloroplasts.

The green algae are of considerable interest because of the possibility that from them the higher plants may have arisen, and because of the diversity of forms which are found within in the group.

The protoplasts of the green algae are with very few exceptions enclosed within a rigid wall composed of two layers, an inner made up wholly or largley of cellulose and an outer layer of pectin (see **Carbohydrates**). Within the protoplast of the cell is located one or more conspicuous chloroplasts containing pigments approximately like those occurring in the plastids of higher plants. The chloroplasts of any single genus are usually very constant in appearance but in the different genera remarkable diversity of size and shape exists. The primitive form seems to be the massive cup-shaped type such as occurs in many lower Chlorophyta. The nucleus of all green algae is a definitely organized body possessing a nuclear membrane, one or more nucleoli, nuclear sap (karolymph), and chromosomes. Many unicellular forms have one or more cilia which persist throughout their existence. The reproductive cells of most green algae have cilia. See **Cilium.**

Reproduction in green algae takes place in various ways. One is a strictly vegetative process in which the colony, or filament, of cells is broken up by various external agents, after which each fragment becomes a new colony, or filament.

Asexual reproduction is by means of cell division (nuclear division followed by division of the cytoplasm) or by zoöspore formation. These zoöspores are generally formed from the protoplast of any vegetative cell, and may appear singly or in numbers by division of a single protoplast. They are expelled from the cell in a manner as yet unknown and are frequently enclosed in delicate vesicle at the time of expulsion. The zoöspores are naked bodies, having no cell wall, and possessing apical cilia. *Vaucheria* is an exception, for its zoöspores are covered with cilia. After periods of motility varying from a few minutes to many hours, zoöspores become quiescent, withdraw their cilia, secrete a cell wall and develop to new colonies or organisms like the parent form.

Finally, many green algae possess a sexual reproduction which is often very complicated. In sexual reproduction there are formed two sets of reproductive bodies known as gametes, which fuse in pairs to form a zygote. From the zygote a new plant develops.

The great diversity of form found in the green algae makes them interesting plants to examine for lines of evolution which have produced the many forms existent today. Several different lines have been found. One of these includes many of the forms which remain motile throughout their existence. A relatively simple unicellular organism is the starting point for such a line. *Chlamydomonas* is of this type. The plant is a small spherical or oval cell enclosed in a definite cell wall and having two apical cilia. Within the protoplast there is a single massive chloroplast which has a shape like that of a soft rubber ball pushed deeply in on one side. Within the hollow of the plastid the single distinct nucleus is located. At the apex of the cell near the origin of the two cilia there is a minute red body called an eye-spot. This is a light-sensitive organ which when stimulated causes the organism to move toward or away from the light source. In the apical end there is a pair of contractile vacuoles. *Chlamydomonas* reproduces asexually by means of zoöspores. These are formed by divisions of the protoplast to form 2, 4 or 8 daughter protoplasts, contained within the wall of the original cell. This wall softens and liberates these naked cells. At once they become small replicas of the parent cell. They soon grow to the size of the original cell and again form a new group of zoöspores. In a suitable environment this is a very rapid method of reproduction. *Chlamydomonas* also reproduces sexually.

In sexual reproduction the protoplast of the cell divides to form motile bodies called gametes, which are quite like zoöspores, but smaller. On liberation from the parent cell wall, gametes from different cells unite in pairs, and form zygotes which develop into zygospores. A zygospore is a resistant spore with a thick wall which enables the organism to survive periods of adverse conditions. When favorable conditions return, the contents of the zygospore divide to form zoöspores, which behave as do similar spores from motile cells.

The primitive character of *Chlamydomonas* is found in its contractile vacuoles, its eye-spot, its single massive plastid, and its cilia. From Flagellates it differs only in having a definite cell wall. Comparing other motile green algae with *Chlamydomonas* makes it possible to discover an interesting series of species of increasing complexity. First in this series is *Gonium sociale*, with colonies of four cells, and *Gonium pectorale*, with sixteen cells, held together loosely in a gelatinous matrix. All the cells of a colony are alike and any cell may form either zoöspores or gametes. All gametes are alike, but fusion occurs between gametes from different colonies. Next in the line of increasing complexity is *Pandorina*, in which a spherical colony is formed. In this genus the gametes are slightly different in size and behavior, some being small and active, while others, slightly larger, are more sluggish. This is an indication of a differentiation of sex.

Still further advance is shown by *Eudorina*, in which a colony is composed usually of 32 cells located in the peripheral portion of the gelatinous matrix. Each cell of the colony is like a *Chlamydomonas* cell and each is capable of reproducing asexually. But in sexual reproduction a very obvious difference in sexes is apparent. Some colonies are definitely female, the cells enlarging slightly and functioning as eggs. In other colonies, each cell divides to form 64 minute biciliate sperm. Fusion between an egg and a sperm produces an oöspore which gives rise to a new colony. In *Pandorina* then a very obvious distinction between sexes has appeared, but the vegetative cells remain alike. In *Pleodorina* each colony is composed of small, purely vegetative cells and larger reproductive cells. The smaller vegetative cells are formed in the anterior end of the colony. In *Volvox* we find the highest degree of differentiation exhibited in this line of motile algae. In this plant the number of cells in a colony is very great, in some species there being as many as 25,000. Of these cells only a few are reproductive, while thousands remain vegetative. Many *Volvox* colonies are so large as to be readily visible to the unaided human eye. In these the many cells form a single layer embedded in the gelatinous matrix. The center of the colony is either water or a thin gelatinous substance. The cells of the colony have the same structure as *Chlamydomonas* cells. They are joined together by fine protoplasmic strands. The beating of the cilia causes the large colony to roll rapidly about in the water, making it a fascinating object to watch. In asexual reproduction certain cells of the colony enlarge and move to the central region. There they lose their cilia, after which they divide rapidly to form new colonies which remain for some time within the parent. Often a single colony will contain a dozen or more of these small colonies. The latter are liberated by the disintegration of the parent colony. In sexual reproduction, cells in the posterior region of the colony differentiate. Some lose their cilia and become very large; these are eggs. Other cells, either in the same or different colonies, divide many times to form large numbers of minute biciliate sperm. These swim to the eggs. A single sperm enters an egg, its nucleus fusing with that of the egg. As a result a zygote is formed. This secretes around itself a thick wall and becomes an oöspore, capable of enduring protracted periods of unfavorable conditions. With the return of favorable conditions the thick wall of the oöspore breaks, the protoplast emerges and divides rapidly, forming a new colony. *Volvox* represents the climax reached in this line of evolution. There is not only a very great increase in the number of cells forming a colony, but also a distinct separation of vegetative and reproductive cells. The reproductive cells are of two kinds, eggs and sperm. But every cell of the plant retains the primitive character of the individual cell.

It is possible to build up other evolutionary series of green algae in which the vegetative cells are non-motile. *Chlamydomonas* often assumes a non-motile condition; the cells become embedded in a copious gelatinous matrix and lose their cilia. This condition is known as the palmella stage. *Tetraspora* is an alga which has the appearance of the palmelloid stage of *Chlamydomonas*. See Fig. 3. Another genus, *Palmella*, normally exists as a shapeless colony of cells held together in a gelatinous matrix. Cilia are lacking, but may be developed by any cell in the colony. A ciliated cell escapes and swims about freely for a time, then settles down and divides to form a new colony. Asexual zoöspores are formed in *Palmella* and also isogametes, that is, gametes of equal size. *Palmella* shows the beginning of a non-motile habit, with a restriction of the motile stages to the reproductive cells. In *Geminella* the amorphous habit of the colony is lost; divisions take place in such a way that the resulting cells tend to exist in a single series, the individual cells being held together only by the gelatinous matrix around them. In *Ulothrix* further advance is made. See Fig. 4. In this plant the protoplast of the cell divides within the wall of the cell. But it does not escape therefrom; instead cross-walls are formed between daughter protoplasts which remain permanently joined. Repeated divisions in a single direction result in the formation of a long unbranched filament of cells. Asexual reproduction in *Ulothrix* is by zoöspores. These are very similar to the cells of *Chlamydomonas*, but each has 4 cilia. From each zoöspore new filaments are formed directly. Sexual reproduction is by biciliate gametes which are formed in the usual manner. Two gametes fuse to form a zygote, which develops a thick wall, becoming a zygospore, and on germinating produces zoöspores. In some species of *Ulothrix* the gametes are alike, while in others slight differences in size of the gametes produced from different cells indicate the beginning of sex differentiation.

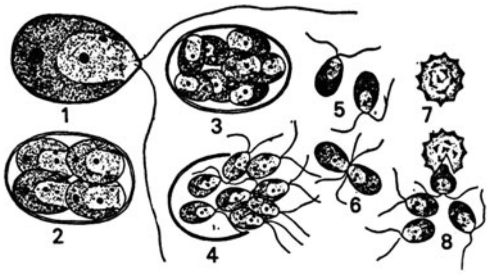


Fig. 3. *Chlamydomonas*: (1) mature cell; (2) four young cells formed in asexual reproduction; (3) eight gametes formed by the division of one cell; (4) gametes escaping from old cell wall; (5) free-swimming gametes; (6) two gametes conjugating; (7) resting zygote; (8) four cells formed from the zygote.

Other genera of algae, related to *Ulothrix*, show greater differentiation in the vegetative cells. Branching occurs in many; certain cells produce the zoöspores or gametes. Other genera, notably the marine *Ulva*, the Sea Lettuce, have cell divisions in two planes, so that extensive membranes are formed, often two meters or more in length.

The highest stage of development in this series is found in *Coleochaete*. This alga appears as small disks epiphytic on other aquatic plants. It consists of branching filaments which in some species grow out to form a flat shield-like thallus. Asexual reproduction is by biciliate zoöspores, which are formed singly from the protoplast of any cell. In its sexual reproduction, *Coleochaete* is especially interesting, exhibiting an advanced degree of differentiation of sex cells called oögamy. The eggs are formed in special cells at the tips of certain filaments. By continued growth, the vegetative cells form a layer enclosing the oögonium. From the oögonium a projection called the trichogyne grows out; in some species it remains short, but in one at least it becomes long and slender. The biciliate sperm are formed singly in special cells known as antheridia. A single sperm enters the egg through the trichogyne or papilla, through a soft place which forms in the wall. Following fertilization, or the union of egg and sperm, the fertilized egg enlarges greatly and secretes around itself a thick wall, in which condition it remains very resistant to external changes. On germination its contents divide to produce 16 or 32 cells, each of which produces a zoöspore. These give rise to new plants. In *Coleochaete* we have a highly developed end product to an advancing degree of differentiation of cells and development of sex.

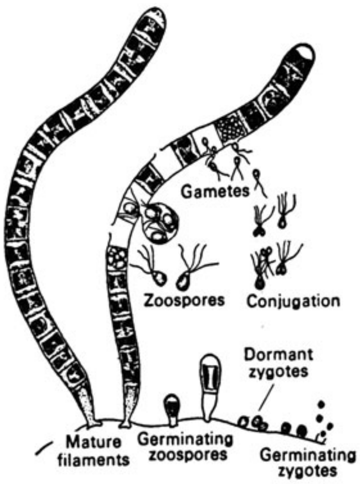


Fig. 4. *Ulothrix*. Stages in development.

Many other equally interesting lines of evolution can be found in green algae. One leads to *Vaucheria*, a branching filamentous plant in which cross walls are not formed, nuclear divisions occurring until an extensive multinucleate filament is formed. Many marine relatives of *Vaucheria*, especially abundant in tropical seas, have elaborate bodies, often thickly encrusted with lime. In the genus *Caulerpa* the thallus of some species appears to be differentiated into leaves, stems and roots; however, no differentiation of tissue occurs, the whole structure being essentially filamentous.

Another group of algae, found only in fresh water, is the Conjugales. This includes the frequently observed *Spirogyra*, a genus having spiral chloroplasts. See Fig. 5. Sexual reproduction in *Spirogyra*, as in all the Conjugales, is by a process called conjugation. When the cells of two filaments are to conjugate they come into contact one with another. From the cells of one filament projections grow outward to meet similar projections from the cells of the other filament. These projections meet, and the walls at their tips are absorbed, leaving a tube connecting two cells of opposite filaments. Through this tube the protoplast of one cell moves into the opposite cell, where fusion occurs. Following this the zygote thus formed surrounds itself with a thick wall. This zygospore is extremely resistant to unfavorable external conditions, such as drought. After a prolonged rest period, if environmental conditions are favorable, the wall of the zygospore breaks, and a tube grows out from the protoplast within. By divisions, this tube becomes a new filament. Observation of conjugating filaments of *Spirogyra* reveal two facts; first, that when one cell of a filament conjugates with a cell of another filament, commonly all the cells of the first filament are conjugating with those of the other; second, that movement is largely from one filament to the other, so that at the end of the process one filament is com-

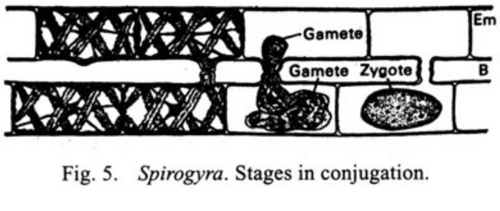


Fig. 5. *Spirogyra*. Stages in conjugation.

posed of empty cell walls and the other filled with zygospores. This may be conceived as a sexual condition, the empty cells having been male and the cells in which the zygospores formed, female. Many related algae in the Conjugales show no indication of sexuality, the protoplasts uniting in the tube joining the two cells. No ciliated cells of any sort occur in this order.

Related to *Spirogyra* is a family containing many species whose cells are very symmetrical and beautiful objects. This is the Desmidiaceae, a family whose members commonly occur as single cells composed of two symmetrical halves. In many species the halves are distinctly indicated by a deep constriction which leaves only a slender isthmus connecting them. Conjugation in the Desmids is essentially like that in the other Conjugales.

Except for their botanical interest and for the fact that they provide food for many animals, the green algae are of little significance to man. Recently, however, investigations have been conducted to determine the feasibility of raising green algae as human food. One of the most promising of these algae is Chlorella, and some of its possibilities in food production are described under **Chlorella.**

The brown algae, or Phaeophyta, are distinguished from other algae by the presence of the brown pigment fucoxanthin, which masks the chlorophyll present and which imparts to them a brown color. Nearly all are marine plants. In this group there are no very simple primitive forms, comparable to *Chlamydomonas* of the greens. The simplest forms, such as *Ectocarpus*, are branched filamentous plants.

Other brown algae are very large, with tough bodies of many often complex shapes. They are most numerous on rocky coasts, where they grow attached to the rocks and are often exposed to the severest pounding of the tides. Brown algae are found in all regions, but are especially abundant in the cold temperature and arctic waters. One genus, *Sargassum*, occurs in great abundance floating in the Atlantic Ocean, forming the Sargasso Sea, through which Columbus passed so slowly on his voyage of discovery. Probably these plants are carried by ocean currents from the shores where they grow to the Sargasso Sea, where they float endlessly. The brown algae include the largest plants of this division of the plant kingdom, many of the kelps growing 30–40 feet (9 to 12 meters) long, and some of the plant forms of the Pacific Ocean attaining lengths of a hundred feet and more.

All brown algae are multicellular plants. Each cell contains a single distinct nucleus and several chloroplasts which contain chlorophyll, carotene, xanthophyll and fucoxanthin. The presence of the latter hides the other pigments. Photosynthesis in brown algae, as in other plants, results in the formation of sugar. This sugar however is changed into a compound, laminarin, instead of starch, which is never found in this group of algae.

The Phaeophyta are divided into several groups. These show interesting differences in their life histories. In most of them there is a very definite alternation of generations. Plants of the sexual generation form gametes; the asexual generation forms zoöspores. The motile cells of the brown algae are quite distinct from those of the green algae, having 2 unequal cilia which are attached laterally and extend in opposite directions.

One of the orders of this class is the Ectocarpales, of which *Ectocarpus* is a common form. See Fig. 6. It occurs as soft brown tufts growing on other larger algae or on stones or woodwork in the water. Each plant is composed of slender branching filaments. In this genus there is a definite alternation of two generations which are indistinguishable in the vegetative condition. Plants of the sexual generation are haploid, that is, have the reduced number of chromosomes and bear gametangia. These are elongated structures composed of many small, cubical cells. Each of these cells forms a single gamete. The plants of this generation and the gametangia they bear all look exactly alike. The gametes which they produce also look very much alike. But in behaviour they are different. Some are sluggish, moving but little, while others swim actively and are attracted to the sluggish ones from other plants. The active gametes are males, the others females; sometimes the female gametes are slightly larger than the males. One male gamete fuses with a female, forming a zygote. The zygote always forms an asexual plant which is diploid. There are two kinds of asexual plants, which look exactly alike, and also like a sexual plant. Each forms zoösporangia. One form of zoösporangium, called a plurilocular zoösporangium, is composed of many small cubical cells, each of which forms a single zoöspore. These zoöspores are liberated, swim about for a time, sink to the bottom and settle against any solid substratum and give rise to new asexual plants of the same type as that producing them. The other type of zoösporangium, often found on the same plant as the first, is composed of a single, usually much-enlarged, cell. The protoplast of this cell divides many times and gives rise to several zoöspores which are haploid. There fore reduction division takes place in the sporangia, which produce these zoöspores. These zoöspores produce sexual plants. In some species of *Ectocarpus* the alternation of generations is not as regular as that described.

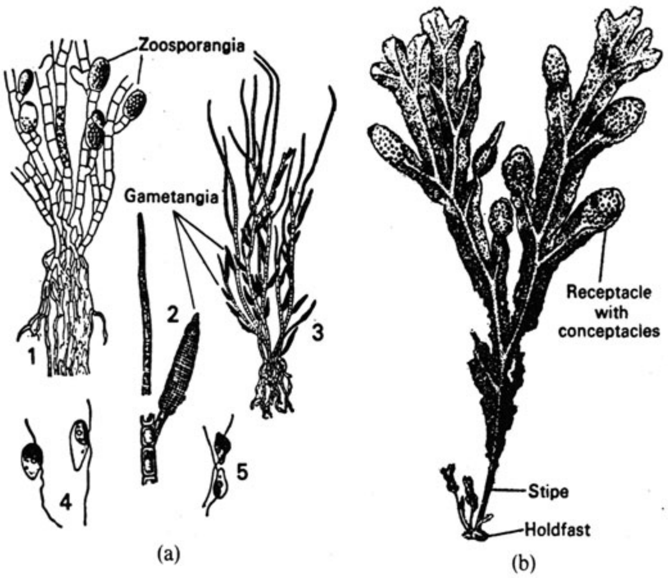


Fig. 6. (a) *Ectocarpus*. Stages in development: (1) sporophyte plant; (2,3) gametophyte plants; (4) two zoöspores; (5) conjugation of gametes to form zygotes. (*Holman and Robbins, "Textbook of General Botany," Wiley*). (b) *Fucus*. Mature plant. No bladders are shown. ("*Morphologie und Biologie der Algen*," *Gustav Fischer, Jena*)

Another order of brown algae is the Cutleriales, of which the genus *Cutleria* is a well-known example. In *Cutleria* the two generations have a very different appearance, the sexual plants being much branched, and several inches tall, while the asexual plants are small lobed thalli growing prostrate on the substratum. So different are these two generations that they are often mistaken for different plants. The sexual plants are of two kinds, male and female, which differ very little. The sex organs are borne on the surface of the thallus. The male gametangia are elongate structures borne on branching filaments; the female gametangia are stouter and composed of few cells. The male gametes are minute and biciliate, the female, also biciliate, are many times larger. Many male gametes swim to a single female, one fuses with it, producing a zygote. The zygote develops into an asexual plant. The zoöspores, formed from small sac-like zoösporangia, which develop in large numbers on the upper surface of the thallus, are biciliate and very similar to gametes. They form sexual plants. In this order there is a very distinct difference in the two gametes, male and female, and a striking difference between sexual and asexual plants.

A third order is the Laminariales, which includes the largest algae known, commonly known as kelps. Some of them have a very striking appearance. *Postelsia*, for example, has a plant body composed of an erect stiff stalk often several inches in diameter. From its base many thick root-like outgrowths spread out and fix the plant firmly to the substratum. At the top of the stalk, long spreading branches are found. When the plant is seen growing in water it has much the appearance of a palm tree, whence the common name, sea palm. Another plant in this order is *Chorda filum*. Its thallus is a tough cord-like object 3–8 feet (0.9 to 2.4 meters) long and about $\frac{1}{4}$ inch (0.6 centimeters) in diameter. It looks very much like a coarse round leather shoe-string. A very common genus is *Laminaria*, which has many species of various forms, mostly large. Some of them consist of long cord-like stalks which bear at their upper end a broad flat expansion often 6–12 feet (1.8 to 3.6 meters) long and 8–15 inches (20.3 to 38.1 centimeters) broad. Colloquially these are known as devil's aprons. These large plants are the asexual generation and so are diploid. The zoösporangia are formed in immence numbers on the surface of the thallus. Each zoösporangium is a cylindrical object which produces many small biciliate zoöspores. These zoöspores swim down to the sea bed, where they develop into haploid or sexual plants. The latter are minute, usually consisting of a few cells which form a branching filament. Some of the plants are male, others female. In the female plant, any cell may become a sexual cell; often the plant is only a single cell. This sexual cell is an oögonium and forms a large non-motile egg which remains in the parent plant. Any cell of the male plant may become sexual, producing minute biciliate sperms which swim to the egg and fuse with it, forming a zygote. The latter at once develops into an asexual plant. In this order, there is also a distinct alternation of generations, but the sexual generation contains the small plants, the asexual usually very large plants. The sexual cells are distinctly different: the large non-motile egg and the small biciliate sperm.

A fourth order of brown algae is the Fucales. Members of this order are tough, much-branched plants which are particularly abundant between the tide levels on rocky shores in regions where the water is cold. *Fucus*, the common rockweed or bladder wrack, is a common and well-known plant. In this, as in all members of this order, there is no asexual reproduction. Therefore no distinct alternation of generations can occur. A *Fucus* plant consists of a tough dichotomously branched frond, which is attached to the rock on which it grows by a disk-shaped holdfast. In many species hollow bladders develop along the frond and serve to bring the plant into an erect position at high tide. At the tips of the branches of the thallus the reproductive bodies, receptacles, are formed. In some species these tips are swollen to form hollow bladders, in others they are flat and little differentiated from the rest of the thallus. The reproductive cells are formed in spherical cavities which are connected with the surface by small pores. Each cavity is called a conceptacle. Numerous branching filaments rise from the lower part of the conceptacle wall. Branches of these filaments bear the sexual organs. In some species the two sexes are borne in the same receptacle, in others they occur on different plants. The male sex organs or antheridia are oval sacs. The protoplast of each sac divides to form 64 cells, each of which becomes a laterally biciliate sperm. When mature these antheridia are extruded through the ostiole or opening of the conceptable into the water. There the wall of the antheridium bursts, liberating the sperms. Each oögonium consists of a single cell. Its protoplast divides to form eight eggs. These also are extruded from the conceptacle, while still within the wall of the oögonium, and freed by the bursting of the same. Both types of conceptacle secrete a gelatinous matrix which surrounds the oögonia or antheridia and aids in bringing them to the surface of the receptacle. This matrix is squeezed out of the conceptacles by the partial drying of the frond at low tide. Each egg is a very large non-motile cell. Thousands of sperms are attracted to each egg and swim about it, causing it to revolve rapidly. Finally one sperm gains entrance to the egg and fertilizes it. The other sperms immediately swim away. The fertilized egg settles to the bottom, attaches itself to the substratum and at once starts to develop into a new plant. In the Fucales there is no alternation of generations. The gametes of the Fucales are very distinct, one being a large non-motile egg, the other a very small swimming sperm. One is tempted to arrange the various orders of brown algae in a series showing the way in which each may have evolved. However, such evolutionary relationships are purely speculative and not supported by any real evidence. It is impossible to trace the ancestry of the brown algae back to any simple ancestor, since no simple forms of brown algae are known.

The economic importance of the brown algae, while slight, is much greater than that of the green algae. Large quantities of these plants are gathered and used for fertilizer, wherever agriculture is carried on near the coast. From the ash produced by burning the larger forms, the kelps and Fucales, iodine and also potassium are obtained. In the Orient and in some of the north Atlantic islands, some of the brown algae are used as food, both for human beings and for livestock.

The red algae or Rhodophyta form a very large group of plants, nearly all marine, of small to medium size. See Fig. 7. They are particularly abundant in warm coastal waters and are often plants of great beauty and extremely delicate habit. The red color to which they owe their name is caused by phycoerythrin, a red pigment which is present with the common chlorophyll-carotene group of pigments. These pigments are present in definite bodies or plastids, and not diffused through the protoplasts, as in the blue-green algae.

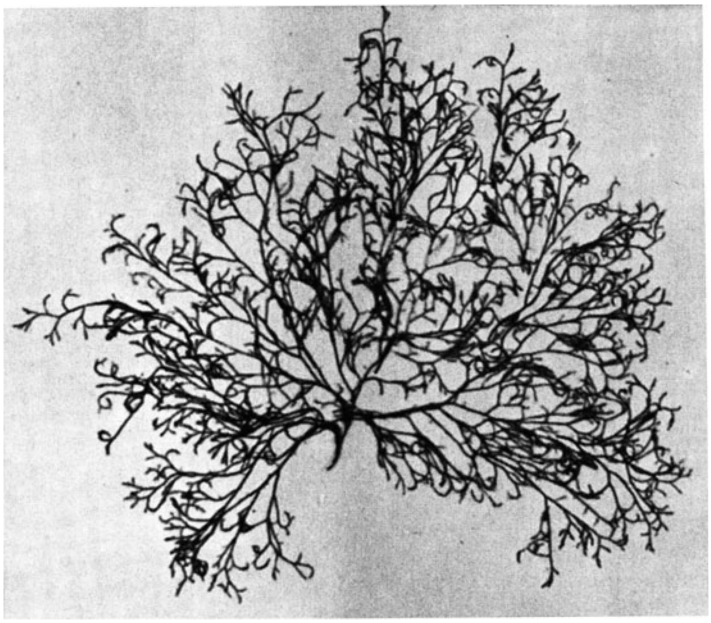


Fig. 7. Red algae. (*A. M. Winchester*)

The forms of red algae are numerous. In many species the thallus is an extremely delicate filament. In other species the thallus is a tough, branched body 6–15 inches (15.2 to 38.1 centimeters) long. Others are flat membranes which may be a single cell in thickness or may be many cells thick. Some are thickly covered with a calcareous deposit, so that they are hard and stony, resembling corals. No motile reproductive cells are produced by members of this group. In sexual reproduction there is always a large female cell which is fertilized by a small male cell. This sexual reproduction is a rather complicated process. The antheridia are single-celled bodies; in some species the whole cell is liberated, in others the protoplast of the antheridium is freed. In either case the male cell floats in the water, carried only by the current. The female reproductive organ is known as the procarp. In simpler forms this consists of a swollen basal portion called a carpogonium and a long slender portion called a trichogyne. Chance brings the male cell to the surface of the trichogyne, against which it sticks. The wall of the trichogyne is dissolved, allowing the nucleus of the male cell to enter the trichogyne. This nucleus passes down the trichogyne and enters the carpogonium, where it fuses with the female nucleus. From the fertilized carpogonium asexual spores called carpospores are formed, usually at the ends of branches which grow out from the carpogonium or from cells which are formed from those surrounding the carpogonium. Into these carpospores, nuclei from the carpogonium pass. The carpospores of simpler red algae at once produce sexual plants. In most red algae, however, they produce asexual plants which may be identical with the sexual plants in appearance. These asexual plants bear reproductive cells called tetraspores, because 4 of them are borne in a single sporangium. Each tetraspore gives rise to a sexual plant. So in the majority of red algae there is a distinct alternation of generations.

Of the many species of red algae very few are of any importance. In northern waters of both coasts of the Atlantic, dulse, *Rhodymenia palmata*, is found. It is gathered, cleaned more or less, and dried. It is then sold as a food or a relish. Species of *Porphyra*, often called laver, are also eaten, especially by Oriental people. Irish moss, or carrageen, which is *Chondrus crispus*, is another red alga which is gathered for food. See Fig. 8. It is a small much-branched plant, commonly dark red in color and with a beautiful iridescent surface. The plants are gathered, thoroughly cleaned and dried. Drying bleaches them to a creamy white color. When thoroughly dry they are bagged and sold. The powdered plant is commonly boiled in milk, flavored and sweetened, and allowed to cool. It forms a firm smooth gel known as blanc-mange. From species of red algae growing in the Pacific Ocean, agar is obtained.

The origin of the red algae and their relationships with other algae is a matter of considerable speculation. A few trace them from the blue-green algae, finding in some of the more primitive red algae "connect-

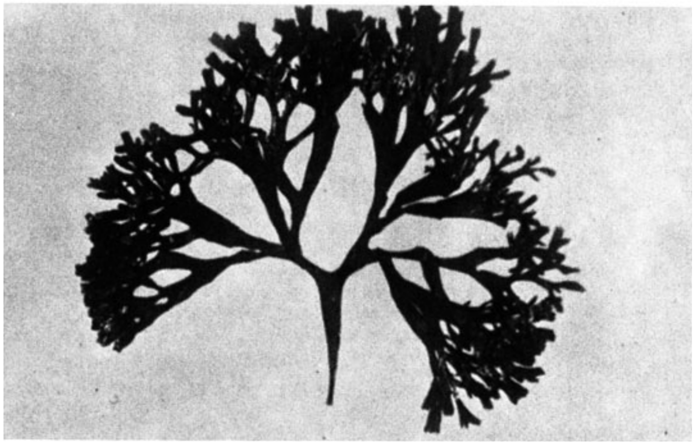


Fig. 8. *Chondrus crispus*. Frond. (*Photomicrograph by B. J. Ford; copyright*)

ing links" which support this view. There are no ciliated reproductive cells in either of the two groups. But their distinct well-developed nucleus, their plastids, and their complex reproductive process set the red algae off very clearly. In view of these facts, it is perhaps more logical to derive the red algae from the green, using a form like *Coleochaete*.

Additional Reading

Cowen, R.: "Parasite Power," *Science News*, 200 (September 29, 1990).

Fogg, G. F., et al.: "The Blue-Green Algae," Academic Press, London, 1973.

Humm, H. J., and S. B. Wicks: "Introduction and Guide to the Marine Bluegreen Algae," Wiley, New York, 1980.

Irvine, D. E. G., and J. H. Price, Ed.: "Modern Approaches to the Taxonomy of Red and Brown Algae," Academic Press, London, 1978.

Roberts, D. A., and C. W. Boothroyd: "Fundamentals of Plant Pathology," Freeman, Salt Lake City, Utah, 1984.

Round, F. E.: "The Ecology of Algae," Cambridge University Press, New York, 1981.

Shubert, L. E., Ed.: "Algae as Ecological Indicators," Academic Press, London, 1984.

ALGEBRA. 1. A generalization of arithmetic involving the study of relations between numbers represented by symbols and obtained by the operations of addition, subtraction, multiplication, division, raising to a power, and extracting a root. If only the first four operations are performed a finite number of times the resulting numbers, functions, or equations are rational; more precisely, rational integral or polynomial and rational fractional. When fractional powers and extraction of roots are also considered the quantities are irrational. A common problem in algebra is that of finding the value of an unknown satisfying an equation written in terms of the operations just stated. Other kinds of numbers, functions, or equations are not algebraic but transcendental.

2. An algebra over a field is defined as follows:

Let K be any field (in physical practice, usually a complex field) whose elements we shall call scalars. Let the set γ_{ijk}, with $i, j, k = 1, 2, \ldots, n$, be any n^3 elements of K. Then the set of all ordered n-tuples $(x_1, x_2, \ldots, x_n)$ of elements of K is called an algebra over K if addition, multiplication and scalar multiplication are defined thus:

$$(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$$

$$(x_1, x_2, \ldots, x_n)(y_1, y_2, \ldots, y_n) = (\gamma_{ij1}x_iy_j, \gamma_{ij2}x_iy_j, \ldots, \gamma_{ijn}x_iy_j)$$

$$\lambda(x_1, x_2, \ldots, x_n) = (\lambda x_1, \lambda x_2, \ldots, \lambda x_n)$$

where λ is a scalar and we have used the summation convention. Setting

$$l_1 = (1, 0, \ldots, 0), l_2 = (0, 1, \ldots, 0), \ldots, l_n = (0, 0, \ldots, 1)$$

we have $(x_1, x_2, \ldots, x_n) = x_il_i$, and $l_il_j = \gamma_{ijk}l_k$. The l_i form a basis of the algebra and the γ_{ijk} are its structure constants.

3. An algebra of subsets of a set X is a class of subsets of X which contains the complement of each of its members and the union of any two of its members (or the intersection of any two of its members). An algebra of subsets is a Boolean algebra relative to the operations of union and intersection. (For definitions of union and intersection as used here, see **Boolean Algebra.**)

4. The algebra of a group is that of polynomials, with coefficients in a field K, of the elements $a_1, a_2, \ldots, a_n$ of a finite group G, where if

$$x = k_1a_1 + k_2a_2 + \ldots + k_na_n$$

$$y = k'_1a_1 = k'_2a_2 + \ldots + k'_na_n$$

then

$$x + y = (k_1 + k'_1)a_1 + \ldots + (k_n + k'_n)a_n$$

and

$$xy = l_1a_1 + l_2a_2 + \ldots + l_na_n$$

where x and y have been multiplied together as polynomials, the product a_ia_j being determined by the law of multiplication of the group. See also **Algebraic Equations.**

ALGEBRAIC EQUATIONS. An equation, or set of simultaneous equations, in which the unknowns occur as rational functions only. Hence the equations are expressible by equating polynomials to zero. Here the case of a single equation in one unknown is considered:

$$a_0P(x) \equiv a_0x^n + a_1x^{n-1} + \cdots + a_n = 0, a_0 \neq 0$$

where the a_i do not depend upon x, and are called the coefficients of the equation, and n is the degree. This equation is equivalent to

$$P(x) \equiv x^n - c_1x^{n-1} + \cdots + (-1)^nC_n = 0$$

$$c_i = (-1)^ia_i/a_0$$

where the c_i are the elementary symmetric functions of the roots x_i. See also **Symmetric Function.**

The *remainder theorem* states that if $a_0P(x)$ is divided by $x - r$, the remainder is $a_0P(r)$:

$$a_0P(x) \equiv (x - r)Q(x) + a_0P(r)$$

The *factor theorem* is a corollary and states that if r is a root of $P(x) = 0$, then $x - r$ divides $P(x)$. The "fundamental theorem of algebra" states that every algebraic equation has a root, real or complex. These theorems imply that $P(x)$ can be factored completely:

$$P(x) = (x - x_1)(x - x_2) \cdots (x - x_n)$$

each x_i being a root. If $x_i = x_j$ for some $i \neq j$, then $x_i = x_j$ is a double root; if $x_i = x_j = x_k$, a triple root, ... Counting a root x_i of multiplicity m as being m coincident roots, one says that an algebraic equation of degree n has exactly n roots (neither more nor less).

A Taylor series expansion gives

$$P(x - r) = P(r) + (x - r)P'(r) + \cdots + (x - r)^nP^{(n)}(r)/n!$$

Hence r is a root of multiplicity m if and only if

$$0 = P(r) = P'(r) = \cdots = P^{(m-1)}(r) \neq P^{(m)}(r)$$

hence r satisfies the derived equations

$$P^{(i)}(x) = 0, \quad i = 0, 1, \ldots, m - 1$$

In general, setting $y = x - r$, if

$$b_i = P^{(n-i)}(r)/(n - i)!$$

then any root of

$$y^n + b_1y^{n-1} + \cdots + b_n = 0$$

is r less than a root x_i, hence the roots are said to have been reduced by r. Repeated synthetic division can be applied to the original equation to obtain the b_i, since $b_n = P(r)$ is the remainder after dividing $P(x)$ by x

$- r$; b_{n-1} that after dividing the quotient by $x - r$; . . . . See also **Synthetic Division.**

Other useful transformations are the following. The roots of

$$a_0y^n - a_1y^{n-1} + \cdots + (-1)^n a_n = 0$$

are the negatives of the roots of the original; those of

$$a_0y^n + a_1\alpha y^{n-1} + \cdots + a_n\alpha^n = 0$$

are α times the roots of the original; those of

$$a_ny^n + a_{n-1}y^{n-1} + \cdots + a_0 = 0$$

are the reciprocals of those of the original. These can be derived by setting $x = -y$, $x = y/\alpha$, and $x = 1/y$.

When the coefficients a_i are integers, then any rational root when expressed as a fraction p/q in lowest terms is such that p divides a_n and q divides a_0 without remainder. In particular, any integral root must be a divisor of a_n. In principle all rational roots can be found exactly, and once any root r is known, all other roots must satisfy $Q(x) = 0$ where $Q = P/(x - r)$. For irrational roots, see **Budan Theorem; Iterative Methods;** and **Newton's Formula for Interpolation.**

While these methods, except for Horner's, apply as well to complex roots as to real, it may be convenient to evaluate $P(z)$ with $z = x + iy$, and write

$$P(z) = R(x, y) + iJ(x, y)$$

after collecting real and pure imaginary terms. Then

$$R(x, y) = J(x, y) = 0$$

are the two simultaneous equations in x and y, and any real solution (x, y) determines a complex solution $z = x + iy$ of $P(z) = 0$. See also **Bernoulli Method.**

ALGEBRAIC GEOMETRY. See **Geometry.**

ALGEBRAIC TOPOLOGY. See **Topology.**

ALGICIDE. A substance, natural or synthetic, used for destroying or controlling algae. The term is also sometimes used to describe chemicals used for controlling aquatic vegetation, although these materials are more properly classified as aquatic herbicides. See **Herbicide.**

ALGIN. A hydrophilic colloidal polysaccharide obtained from several species of brown algae. The term is used both in reference to the pure substance, alginic acid, extracted from the algae and also to the salts of this acid such as sodium or ammonium alginate, in which forms it is used commercially. The alginates currently find a large number of applications in the paint, rubber, pharmaceutical, food, and other industries. See also **Gums and Mucilages.**

ALGOL (β Persei). One of the first variable stars to be recognized as such. The first scientific notice of this variability was made by Montanari in 1670, but it is quite evident that the changes in the light of this star were noticed long before then. The name Algol, which signifies "demon star" probably was suggested because of the star's peculiar behavior. Algol is an eclipsing binary and the first star of this type to be explained. Because of its great brightness, it has been extensively observed with all types of stellar photometers, and the characteristics of its light curve are known. Algol is also a spectroscopic binary and, from the determination of the orbital elements from light variability as well as from the spectroscopic data, the physical characteristics of the component parts may be determined. See also **Binary Stars; Eclipsing Binary;** and **Spectroscopic Binaries.**

ALGONKIAN. A term applied to rock formations of late pre-Cambrian (Proterozoic) date in the Great Lakes region, and to the unit of time represented by these formations. Some American geologists use the term as a synonym for late pre-Cambrian.

ALGORITHM. A term derived from the word *algorism*, which meant the art of computing with Arabic numerals. The term *algorithm* is now used (1) to denote any method of computation, whether algebraic or numerical, or (2) any method of computation consisting of a comparatively small number of steps; the steps to be taken in a preassigned order and usually involving iteration, which are specifically adapted to the solution of a problem of some particular type. The best known is Euclid's algorithm for finding the highest common factor of two given numbers.

In computer terminology, an algorithm is a detailed logical procedure which represents the solution of a particular problem. Most commonly, the term is used to indicate an analysis procedure such as that used for the evaluation of a square root or for sorting a data file. Programming algorithms are widely published in the computer field literature. See also alphabetical index.

Algorithms are particularly important where a formerly analog data collection and processing system has changed over to digital methodologies. This has been a frequent occurrence during the last several years, resulting from digital computerization. An excellent example is found in process instrumentation and control and in industrial automation systems. The conversion requires an appropriate software architecture and the matching of digital algorithms to analog behavior.

Whereas the analog controller continuously sensed the process state and manipulated the actuators, the digital controller must repeatedly sample the state, convert it to a quantized number, use that number to compute control actions, and output those actions. Each of these steps involves its own problems and errors.

Standard digital control texts address the sampling problem but treat the broader control design in terms of neutral, computed parameters not clearly related to the process gains and time constants. Process control depends on standard control algorithms, whose parameters may be set or tuned in terms of known process properties.

The detailed calculation of process control algorithms is beyond the scope of this *Encyclopedia*. The details are well covered in the Bristol article listed.

Additional Reading

Bristol, E. H.: "Basic Control Algorithms," in *Industrial Instruments and Controls Handbook*, 4th Edition (D. M. Considine, Editor), McGraw-Hill, New York, 1993.

Gray, D. M., Karmarkar, N. K., and K. G. Ramakrishnan: "The Karmarkar Algorithm: Adding Wings to Linear Programming," *AT&T Bell Record*, 4–10 (March 1986).

Khosla, P. K., and C. P. Neuman: "Computational Requirements of Customized Newton-Euler Algorithms," *J. of Robotic Systems*, **2**(3), 309–327 (1985).

Neuman, C. P.: "A Robot Dynamics Simulator," in *Standard Handbook of Industrial Automation* (D. M. and G. D. Considine, Eds.), Chapman and Hall, New York, 1986.

Tu, F. C. Y., and J. Y. H. Tsing: "Synthesizing a Digital Algorithm for Optimized Control," *Instrumentation Technology*, 52–56 (May 1979).

ALIASING ERROR (Data Acquisition System). The Shannon-Nyquist sampling theorem states that in order to reconstruct a signal containing frequency components in the spectrum 0 to f_m Hz from sampled data, samples must be taken at a rate of at least $2f_m$ samples per second. Failure to obtain data at this rate converts high-frequency components into low-frequency components. Thus, the reconstructed signal contains low-frequency energy not present in the original signal. Figure 1 illustrates a 5-Hz signal sampled at a 4-samples-per-second rate. The reconstructed signal has a frequency of 1 Hz. An *aliasing error* is an error which can be introduced into sampled data in a digital-data-acquisition system as the result of violation of the basic sampling theorem.

The term *aliasing* results from the interpretation that the high-frequency components take the "alias" of a lower-frequency component. An equivalent term, *folding error,* arises from the interpretation that the frequency spectrum of the signal is folded such that high-frequency components appear in a lower-frequency spectrum.

Sampling a continuous signal is equivalent to modulating the signal with a series of uniformly spaced impulses as indicated in Fig. 2. If a signal containing frequency components up to a frequency f_m is modulated by a carrier of frequency f_s, the frequency spectrum of the resulting sampled signal consists of the original spectrum, plus harmonic spectra centered on the sampling frequency and its harmonics

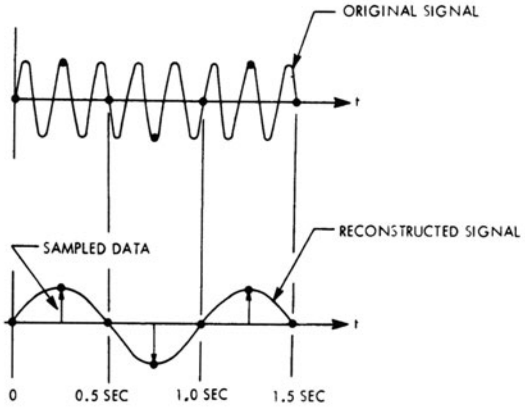


Fig. 1. Example of aliasing error in a sampled-data system.

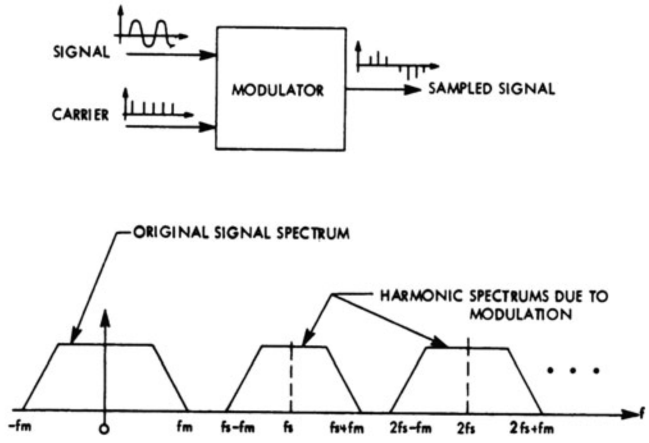


Fig. 2. Frequency-domain sampled spectra.

as shown in Fig. 1. This is the same effect as the modulation of an rf carrier with an audio signal in amplitude-modulated radio, except that there are an infinite number of carrier frequencies because the frequency spectrum of a train of equally spaced impulses (the modulating signal) consists of an infinite number of frequency components at f_s and its harmonics.

Figure 1 shows that if the sample frequency f_s is less than $2f_m$, the lower sideband of the first-harmonic spectrum overlaps the original signal spectrum. Inasmuch as mathematical signal-reconstruction methods are equivalent to low-pass filtering in the signal domain, it is evident that the signal energy from the first-harmonic spectrum which overlaps the original spectrum will be passed by the low-pass filter and, therefore, will affect the nature of the reconstructed signal.

In the foregoing description, the signal was assumed to have no frequency components above the frequency f_m. In practice, signals are not definitively band-limited and high-frequency components are present in any real signal. In the majority of process-control applications, the signals are nominally band-limited by the low-pass characteristics of the process or the signal transducers. In connection with such signals, the sampling theorem frequently is stated so as to require sampling at twice the highest *significant* signal frequency. In practice, selection of the sampling frequency depends upon the interpretation of the term *significant*. A rule of thumb is that the sampling rate should be 5 to 10 times the highest frequency of interest. Although not technically perfect, the rule is valuable inasmuch as signals are usually band-limited to reject frequencies above those of interest. The 5 or 10 multiplier provides a safety factor to account for the finite rolloff rate of most low-pass systems. Where accurate reconstruction of the sampled signal is needed, an estimate of the aliasing error should be made.

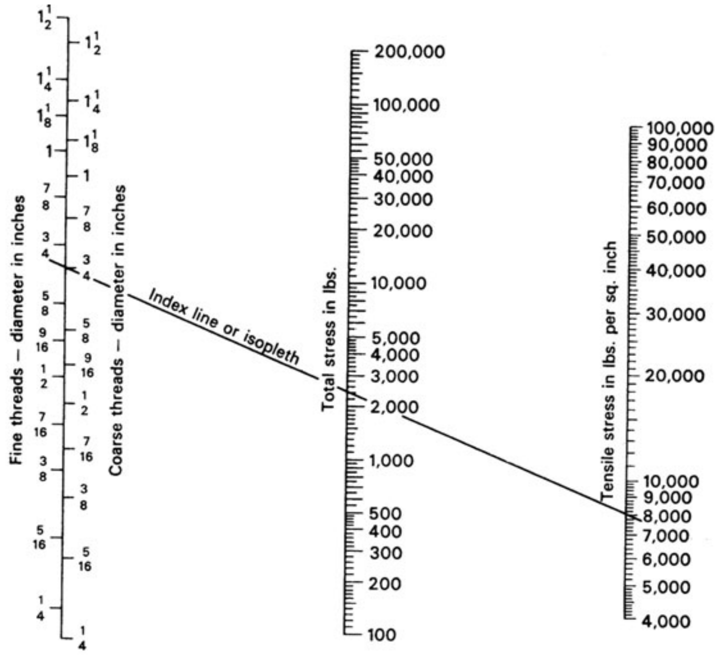
An estimate of the aliasing error can be made by using a graphical technique, based on a frequency-spectrum diagram of the type shown in Fig. 2. The frequency spectrum of the original signal and the spectra of the modulation harmonics are drawn as indicated. At each frequency in the bandpass of the reconstruction filter or its mathematical equivalent, the amplitude of the original frequency components and all contributions from aliased harmonics are summed. This sum is an estimate of the magnitude of this frequency component in the reconstructed signal. The estimate is conservative inasmuch as the true contribution is represented by the vector sum of the components, whereas the estimate is based upon the algebraic sum.

Thomas J. Harrison, International Business Machines Corporation, Boca Raton, Florida.

ALIENATION (Coefficient of). A term occurring in psychology. If the coefficient of correlation between two variables is r, the coefficient of alienation is defined as $(1 - r^2)^{1/2}$.

ALIGNMENT CHART. Nomograms or calculating charts used to represent formulas containing three or more variables. Graduated lines represent the variables; an index line or isopleth passing through points of two scales of known values will intersect the third scale to give the solution of the problem. The nomogram shown will give the solution of the equation $0.785D^2S = T$, where D is the root diameter of an American Standard screw, S is the unit tensile strength, and T is the total stress.

In addition to the accompanying three-parallel-line chart, alignment charts may take other forms, such as straight lines forming a Z shape, or the lines may be curved. The type of chart depends on the form of the mathematical equation represented.



Nomogram for calculating thread diameter and tensile strength and total stress for American Standard screws.

ALIMENTARY TRACT. The structures through which nourishment passes during the process of digestion and elimination. These include the mouth, pharynx, esophagus, stomach, the small intestine, which includes duodenum, jejunum, and ileum, and the large intestine, which includes the cecum, colon, rectum, and anus. See also diagram under **Digestive System (Human).**

ALIMENTATION. In glaciology, the combined processes which serve to increase the mass of a glacier or snowfield; the opposite of ablation. The deposition of snow is the major form of glacial alimentation, but other forms of precipitation, along with sublimation, refreezing of melted water, etc., also contribute. The additional mass produced by alimentation is termed accumulation.

ALIPHATIC COMPOUND. An organic compound that can be regarded as a derivative of methane, CH_4. Most aliphatic compounds are open carbon chains, straight or branched, saturated or unsaturated. Originally, the term was used to denote the higher (fatty) acids of the $C_nH_{2n}O_2$ series. The word is derived from the Greek term for oil. See also **Compound (Chemical); Organic Chemistry.**

ALKALI. A term that was originally applied to the hydroxides and carbonates of sodium and potassium but since has been extended to include the hydroxides and carbonates of the other alkali metals and ammonium. Alkali hydroxides are characterized by ability to form soluble soaps with fatty acids, to restore color to litmus which has been reddened by acids, and to unite with carbon dioxide to form soluble compounds. See also **Acids and Bases.**

ALKALI METALS. The elements of group 1 of the periodic classification. In order of increasing atomic number, they are hydrogen, lithium, sodium, potassium, rubidium, cesium, and francium. With exception of hydrogen which is a gas and which frequently imparts a quality of acidity to its compounds, the other members of the group display rather striking similarities of chemical behavior, all reactive with H_2O to form strongly alkaline solutions. The elements in the group, including hydrogen, are characterized by a valence of one, having one electron in an outer shell available for reaction. Because of their chemical similarities, these elements, along with *ammonium* and sometimes magnesium, are considered the sixth group in classical qualitative chemical analysis separations.

ALKALINE EARTHS. The elements of group 2 of the periodic classification. In order of increasing atomic number, they are beryllium, magnesium, calcium, strontium, barium, and radium. The members of the group display rather striking similarities of chemical behavior, including stable oxides and carbonates, with hydroxides that are less alkaline than those of group 1. The elements of the group are characterized by a valence of two, having two electrons in an outer shell available for reaction. Because of their chemical similarities, these elements are considered the fifth group in classical qualitative chemical analysis separations.

ALKALI ROCKS. Igneous rocks which contain a relatively high amount of alkalis in the form of soda amphiboles, *soda* pyroxenes, or felspathoids, are said to be alkaline, or alkalic. Igneous rocks in which the proportions of both lime and alkalis are high, as combined in the minerals, feldspar, hornblende, and augite, are said to be calcalkalic.

ALKALOIDS. The term, alkaloid, which was first proposed by the pharmacist, W. Meissner, in 1819, and means "alkali-like," is applied to basic, nitrogen-containing compounds of plant origin. Two further qualifications usually are added to this definition: (1) the compounds have complex molecular structures; and (2) manifest significant pharmacological activity. Such compounds occur only in certain genera and families, rarely being universally distributed in larger groups of plants. Many widely distributed bases of plant origin, such as methyltrimethyl- and other open-chain simple alkylamines, the cholines, and the phenylalkylamines, are not classed as alkaloids. Alkaloids usually have a rather complex structure with the nitrogen atom involved in a heterocyclic ring. However, thiamine, a heterocyclic nitrogenous base, is not regarded as an alkaloid mainly because of its almost universal distribution in living matter. Colchicine, on the other hand, is classed as an alkaloid even though it is not basic and its nitrogen atom is not incorporated into a heterocyclic ring. It apparently qualifies as an alkaloid because of its particular pharmacological activity and limited distribution in the plant world.

Over 2000 alkaloids are known and it is estimated that they are present in only 10–15% of all vascular plants. They are rarely found in cryptogamia (exception, ergot alkaloids), gymnosperms, or monocotyledons. They occur abundantly in certain dicotyledons and particularly in the following families: *Apocynaceaae* (dogbane, quebracho, pereiro bark); *Papaveraceae* (poppies, chelidonium); *Papilionaceae* (lupins, butterfly-shaped flowers); *Ranunculaceae* (aconitum, delphinium); *Rubiaceae* (cinchona bark, ipecacuanha); *Rutaceae* (citrus, fagara); and *Solanaceae* (tobacco, deadly nightshade, tomato, potato, thorn apple). Well-characterized alkaloids have been isolated from the roots, seeds, leaves or bark of some 40 plant families. *Papaveraceae* is an unusual family, in that all of its species contain alkaloids.

Brief descriptions in alphabetical order of alkaloids of commercial or medical importance or of societal concern (alkaloid narcotics) are given later in this entry. See also **Amphetamine; Addiction (Drug); Morphine;** and **Pyridine and Derivatives.**

The nomenclature of alkaloids has not been systemized, both because of the complexity of the compounds and for historical reasons. The two commonly used systems classify alkaloids either according to the plant genera in which they occur, or on the basis of similarity of molecular structure. Important classes of alkaloids containing generically related members are the aconitum, cinchona, ephedra, lupin, opium, rauwolfia, senecio, solanum, and strychnos alkaloids. Chemically derived alkaloid names are based upon the skeletal feature which members of a group posses in common. Thus, indole alkaloids (e.g., psilocybin, the active principle of Mexican hallucinogenic mushrooms) contain an indole or modified indole nucleus, and pyrrolidine alkaloids (e.g., hygrine) contain the pyrrolidine ring system. Other examples of this type of classification include the pyridine, quinoline, isoquinoline, imidazole, pyridine-pyrrolidine, and piperidine-pyrrolidine type alkaloids. Several alkaloids are summarized along these general terms in the accompanying table.

The beginning of alkaloid chemistry is usually considered to be 1805 when F. W. Sertürner first isolated morphine. He prepared several salts of morphine and demonstrated that it was the principle responsible for the physiological effect of opium. Alkaloid research has continued to date, but because most likely plant sources have been investigated and because a large number of synthetic drugs serve medical and other needs more effectively, the greatest emphasis has been placed upon the synthetics.

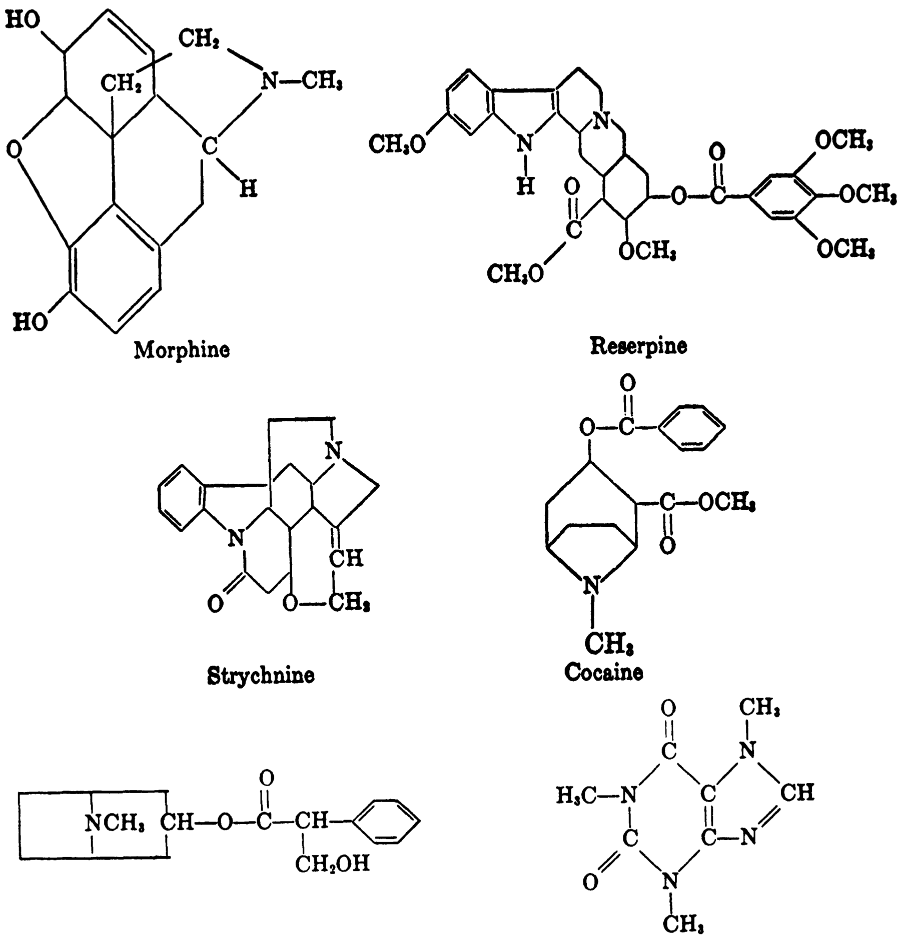
Sometimes, there is confusion between alkaloids and narcotics. It should be stressed that all alkaloids are not narcotics; and all narcotics are not alkaloids. A narcotic has the general definition of a drug which produces sleep or stupor, and also relieves pain. Many alkaloids do not meet these specifications.

The molecular complexity of the alkaloids is demonstrated by the accompanying figure. Alkaloids react as bases to form salts. The salts used especially for crystallization purposes are the hydrochlorides, sulfates, and oxalates which are generally soluble in water or alcohol, insoluble in ether, chloroform, carbon tetrachloride, or amyl alcohol. Alkaloid salts unite with mercury, gold, and platinum chlorides. Free alkaloids lack characteristic color reactions but react with certain reagents, as follows, with (1) iodine in potassium iodide solution, forming chocolate brown precipitate; (2) mercuric iodide in potassium iodide solution (potassium mercuriiodide), forming precipitate; (3) potassium iodobismuthate, forming orange-red precipitate; (4) bromine-saturated concentrated hydrobromic acid forming yellow precipitate; (5) tannic acid, forming precipitate; (6) molybdophosphoric acid, forming precipitate; (7) tungstophosphoric acid, forming precipitate; (8) gold(III) chloride, forming crystalline precipitate of characteristic melting point; (9) platinum(IV) chloride, forming crystalline precipitate of characteristic melting point; (10) picric acid, forming precipitate; (11) perchloric acid, forming precipitate. Many alkaloids form more or less characteristic colors with acids, solutions of acidic salts, etc.

The function of alkaloids in the source plant has not been fully explained. Some authorities simply regard them as by-products of the plant metabolism. Others conceive of alkaloids as reservoirs for protein synthesis; as protective materials discouraging animal or insect attacks; as plant stimulants or regulators in such activities as growth, metabo-

GENERAL CLASSIFICATION OF ALKALOIDS

General Class	Examples
Derivatives of aryl-substituted amines	Adrenaline, amphetamine, ephedrine, phenylephrine tyramine
Derivatives of pyrrole	Carpaine, hygrine, nicotine
Derivatives of imidazole	Pilocarpine
Derivatives of pyridine and piperidine	Anabasine, coniine, ricinine
Containing fusion of two piperidine rings	Isopelletierine, pseudopelletierine
Pyrrole rings fused with other rings	Gelsemine, physostigmine, vasicine, yohimbine
Aporphone alkaloids	Apomorphine corydine, isothebaine
Berberine alkaloids	Berberine, emetine
Bis-benzylisoquinoline alkaloids	Bebeering, trilobine
Cinchona alkaloids	Cinchonine, quinidine, quinine
Cryptopine alkaloids	Cryptopine, protopine
Isoquinoline alkaloids	Anhalidine, pellotine, sarsoline
Lupine alkaloids	Lupanine, sparteine
Morphine and related alkaloids	Codeine, morphine, thebaine
Papaverine alkaloids	Codamine, homolaudanosine, papeverine
Phthalide isoquinoline alkaloids (also known as narcotine alkaloids)	Hydrastine, narceine, narcotine
Quinoline alkaloids	Dictamine, galipoline, lycorine
Tropine alkaloids	Atropine, cocaine, ecgonine, scopolamine, tropine
Other alkaloids	Brucine, sclanidine, strychnine


Morphine

Reserpine

Strychnine

Cocaine

Atropine

Caffeine

Structures of representative alkaloids. The carbon atoms in the rings and the hydrogen atoms attached to them are not designated by letter symbols. However, there is understood to be a carbon atom at each corner (except for the cross-over in the structure of morphine) and each carbon atom has four bonds, so that any bonds not shown or represented by attached groups are joined to hydrogen atoms.

lism, and reproduction; as detoxicating agents, which render harmless (by processes such as methylation, condensation, and ring closure) substances whose accumulation might otherwise cause damage to the plant. While these theories are of interest, it is also of interest to observe that from 85–90% of all plants manage well without the presence of alkaloids in their structures.

Adrenaline®. See Epinephrine later in this entry.

Atropine, also known as daturine, $C_{17}H_{23}NO_3$ (see structural formula in accompanying diagram), white, crystalline substance, optically inactive, but usually contains levorotatory hyoscyamine. Compound is soluble in alcohol, ether, chloroform, and glycerol; slightly soluble in water; mp 114–116°C. Atropine is prepared by extraction from *Datura stramonium*, or synthesized. The compound is toxic and allergenic. Atropine is used in medicine and is an antidote for cholinesterase-inhibiting compounds, such as organophosphorus insecticides and certain nerve gases. Atropine is commonly offered as the sulfate. Atropine is used in connection with the treatment of disturbances of cardiac rhythm and conductance, notably in the therapy of sinus bradycardia and sick sinus syndrome. Atropine is also used in some cases of heart block. In particularly high doses, atropine may induce ventricular tachycardia in an ischemic myocardium. Atropine is frequently one of several components in brand name prescription drugs.

Caffeine, also known as theine, or methyltheobromine, 1,2,7-trimethyl xanthine (see structural formula in accompanying diagram), white, fleecy or long, flexible crystals. Caffeine effloresces in air and commences losing water at 80°C. Soluble in chloroform, slightly soluble in water and alcohol, very slightly soluble in ether, mp 236.8°C, odorless, bitter taste. Solutions are neutral to litmus paper.

Caffeine is derived by extraction of coffee beans, tea leaves, and kola nuts. It is also prepared synthetically. Much of the caffeine of commerce is a by-product of decaffeinized coffee manufacture. See also **Coffee Tree and Coffee.** The compound is purified by a series of recrystallizations. Caffeine finds use in medicine and in soft drinks. Caffeine is also available as the hydrobromide and as sodium benzoate, which is a mixture of caffeine and sodium benzoate, containing 47–50% anhydrous caffeine and 50–53% sodium benzoate. This mixture is more soluble in water than pure caffeine. A number of nonprescription (pain relief) drugs contain caffeine as one of several ingredients. Caffeine is a known cardiac stimulant and in some persons who consume significant amounts, caffeine can produce ventricular premature beats.

Cocaine (also known as methylbenzoylepgonine), $C_{17}H_{21}NO_4$, is a colorless-to-white crystalline substance, usually reduced to powder. Cocaine is soluble in alcohol, chloroform, and ether, slightly soluble in water, giving a solution slightly alkaline to litmus. The hydrochloride is levorotatory, mp 98°C. Cocaine is derived by extraction of the leaves of coca (*Erythroxylon*) with sodium carbonate solution, followed by treatment with dilute acid and extraction with ether. The solvent is evaporated after which the substance is redissolved and subsequently crystallized. Cocaine also is prepared synthetically from the alkaloid ecgonine. Cocaine is *highly toxic* and *habit forming*. While there are some medical uses of cocaine, usage must always be under the direction of a physician. It is classified as a narcotic in most countries. Society's major concern with cocaine is its use (increasing in recent years) as a narcotic.

Cocaine has been known as a very dangerous material since the early 1900s. When use of it as a narcotic increased during the early 1970s, serious misconceptions concerning its "safety" as compared with many other narcotics led and continue to lead to many deaths from its use. For example, health authorities in Dade County, Florida reported in 1977 that over 50% of drug-related overdose deaths were attributable to cocaine (Wetli and Wright, 1979).

Addicts use cocaine intravenously or by snorting the powder. After intravenous injections, coma and respiratory depression can occur rapidly. It has been reported that fatalities associated with snorting usually occur shortly after the abrupt onset of major motor seizures, which may develop within minutes to an hour after several nasal ingestions. Similar results occur if the substance is taken by mouth. Treatment is directed toward ventilatory support and control of seizures—although in many instances a victim may not be discovered in time to prevent death. It is interesting to note that cocaine smugglers, who have placed cocaine-filled condoms in their rectum or alimentary tract, have died (Suarez et al, 1977). The structural formula of cocaine is given in the accompanying diagram.

Codeine, also known as methylmorphine, $C_{18}H_{21}NO_3 \cdot H_2O$, is a colorless white crystalline substance, mp 154.9°C, slightly soluble in water, soluble in alcohol and chloroform, effloresces slowly in dry air. Codeine is derived from opium by extraction or by the methylation of morphine. For medical use, codeine is usually offered as the dichloride, phosphate, or sulfate. Codeine is *habit forming*. Codeine is known to exacerbate *urticaria* (familiarly known as *hives*). Since codeine is incorporated in numerous prescription medicines for headache, heartburn, fatigue, coughing, and relief of aches and pains, persons with a history of urticaria should make this fact known to their physician. Codeine is sometimes used in cases of acute *pericarditis* to relieve severe chest pains in early phases of disease. Codeine is sometimes used in drug therapy of renal (kidney) diseases.

Colchicine, an alkaloid plant hormone, $C_{22}H_{25}NO_6$, is yellow crystalline or powdered, nearly odorless, mp 135–150°C, soluble in water, alcohol, and chloroform, moderately soluble in ether. Solutions are levorotatory and deteriorate under light. The substance is highly toxic (0.02 gram may be fatal if ingested). Colchicine is extracted from the plant *Colchicum autumnale* after which it is crystallized. The compound also has been synthesized. Biologists have used colchicine to induce chromosome doubling in plants. Colchicine finds a number of uses in medicine.

Although colchicine has been known for many years, interest in the drug has been revitalized in recent years as the result of the discovery that it interferes with cell division by destroying the spindle mechanism. The two chromatids which represent one chromosome at the metaphase stage fail to separate and do not migrate to the poles (ends) of the cell. Each chromatid becomes a chromosome in situ. The entire group of new chromosomes now form a resting nucleus and the next cell division reveals twice as many chromosomes as before. The cell has changed from the diploid to the tetraploid condition. Applied to germinating seeds or growing stem tips in concentrations of about 1 gram in 10,000 cubic centimeters of water for 4 or 5 days, colchicine may thus double the chromosome number of many or all of the cells, producing a tetraploid plant or shoot. Offspring from such plants may be wholly tetraploid and breed true. Tetraploid plants are larger than diploid plants and often more valuable. The alkaloid has also been used to double the chromosome number of sterile hybrids produced by crossing widely separated species of plants. Such plants, after colchicine treatment, contain in each cell two complete diploid sets of chromosomes, one from each of the parent species, and become fertile, pure-breeding hybrid species.

In medicine, colchicine is probably best known for its use in connection with the treatment of gout. Acute attacks of gout are characteristically and specifically aborted by colchicine. The response noted after administration of the drug also can be useful in diagnosing gout cases where synovial fluid cannot be aspirated and examined for the presence of typical urate crystals. However, colchicine does not affect the course of acute synovitis in rheumatoid arthritis.

Kaplan (1960) observed that colchicine may produce objective improvement in the periarthritis associated with *sarcoidosis* (presence of noncaseating granulomas in tissue). Colchicine is sometimes used in the treatment of *scleroderma* (deposition of fibrous connective tissues in skin or other organs); it may assist in preventing attacks of Mediterranean fever; and it is sometimes used as part of drug therapy for some renal (kidney) diseases.

Colchicine can cause diarrhea as the result of mucosal damage and it has been established that colchicine interferes with the absorption of vitamin B_{12}.

Emetine, an alkaloid from ipecac, $C_{29}H_{40}O_4N_2$, is a white powder, mp 74°C, with a very bitter taste. The substance is soluble in alcohol and ether, slightly soluble in water. Emetine darkens upon exposure to light. The compound is derived by extraction from the root of *Cephalis ipecacuanha* (ipecac). It is also made synthetically. Medically, ipecac is useful as an emetic (induces vomiting) for emergency use in the treatment of drug overdosage and in certain cases of poisoning. Ipecac should not be administered to persons in an unconscious state. It should be noted that emesis is not the proper treatment in all cases of potential poisoning. It should not be induced when such substances as petroleum distillates, strong alkali, acids, or strychnine are ingested.

Ephedrine, 1-phenyl-2-methylaminopropanol, $C_6H_5CH(OH)CH(NHCH_3)CH_3$, is a white-to-colorless granular substance, unctuous (greasy) to the touch, and hygroscopic. The compound gradually decomposes upon exposure to light. Soluble in water, alcohol, ether, chloroform, and oils, mp 33–40°C, by 255°C, and decomposes above this temperature. Ephedrine is isolated from stems or leaves of *Ephedra*, especially Ma huang (found in China and India). Medically, it is usually offered as the hydrochloride. In the treatment of bronchial asthma, ephedrine is known as a *beta agonist.* Compounds of this type reduce obstruction by activating the enzyme adenylate cyclase. This increases intracellular concentrations of cAMP (cyclic 3′5′-adenosine monophosphate) in bronchial smooth muscle and mast cells. Ephedrine is most useful for the treatment of mild asthma. In severe asthma, ephedrine rarely maintains completely normal airway dynamics over long periods. Ephedrine also has been used in the treatment of cerebral transient ischemic attacks, particularly with patients with vartebrobasilar artery insufficiency who have symptoms associated with relatively low blood pressure, or with postural changes in blood pressure. Ephedrine sulfate also has been used in drug therapy in connection with urticaria (hives).

Epinephrine, a hormone having a benzenoid structure, $C_9H_{13}O_3N$, also called adrenaline. It can be obtained by extraction from the adrenal glands of cattle and also prepared synthetically. Its effect on body metabolism is pronounced, causing an increase in blood pressure and rate of heartbeat. Under normal conditions, its rate of release into the system is constant, but emotional stresses, such as fear or anger rapidly increase the output and result in temporarily heightened metabolic activity. Epinephrine is used for the symptomatic treatment of bronchial asthma and reversible bronchospasm associated with chronic bronchitis and emphysema. The drug acts on both alpha and beta receptor sites. Beta stimulation provides bronchodilator action by relaxing bronchial muscle. Alpha stimulation increases vital capacity by reducing congestion of the bronchial mucosa and by constricting pulmonary vessels.

Epinephrine is also used in the management of anesthetic procedures in connection with noncardiac surgery of patients with active ischemic heart disease. The drug is useful in the treatment of severe urticarial (hives) attacks, especially those accompanied by angioedema.

Epinephrine has numerous effects on intermediary metabolism. Among these are promotion of hepatic glycogenolysis, inhibition of hepatic gluconeognesis, and inhibition of insulin release. The drug also promotes the release of free fatty acids from triglyceride stores in adipose tissues. Epinephrine produces numerous cardiovascular effects. Epinephrine is particularly useful in treating conditions of immediate hypersensitivity—interactions between antigen and antibody. These mechanisms cause attacks of anaphylaxis, hay fever, hives and allergic asthma. Anaphylaxis can occur after bee and wasp stings, venoms, etc. Although the mechanism is not fully understood, epinephrine can play a lifesaving role in the treatment of acute systemic anaphylaxis.

In some instances, epinephrine can be a cause of a blood condition involving the leukocytes and known as neutrophilia. In very rare cases, an intramuscular injection of epinephrine can be a cause of clostridial myonecrosis (gas gangrene).

Heroin, diacetylmorphine $C_{17}H_{17}NO(C_2H_3O_2)_2$, is a white, essentially odorless, crystalline powder with bitter taste, soluble in alcohol, mp 173°C. Heroin is derived by the acetylization of morphine. The substance is highly toxic and is a habit-forming narcotic. One-sixth grain (0.0108 gram) can be fatal. Although emergency facility personnel in some areas during recent years have come to regard heroin overdosage as approaching epidemic statistics, it is nevertheless estimated that the majority of persons with heroin overdose die before reaching a hospital. The initial crisis of an overdose is a severe respiratory depression and sometimes *apnea* (cessation of breathing). In emergency situations, the victim may be ventilated with a self-inflating resuscitative bag with delivery of 100% oxygen. Then, an endotracheal tube attached to a mechnical ventilator may be inserted. Naloxone (*Narcan*®), a narcotic antagonist, then may be administered intraveneously, often with repeated dosages over short intervals, until an improvement is noted in the respiratory rate or sensorial level of the victim. If a victim does not respond, this is usually indication that the situation is not opiate-related, or that other drugs also have been taken. Inasmuch as the antagonizing action of naloxone persists for only a few hours, a heroin overdose patient should be observed in the hospital for an indeterminate period. In heroin overdose cases, pulmonary edema (as the result of altered capillary permeability) may occur. This is directly associated with the overdose and not with subsequent treatment. Aside from the severe overdose situation, use of the drug causes or contributes to a number of ailments. These include chronic renal (kidney) failure and nephrotic syndrome. Septic arthritis caused by *Pseudomonas* and *Serratia* infections, is sometimes found as the result of intravenous heroin abuse. Drug-induced immune platelet destruction also may occur.

Morphine. See separate entry on **Morphine.**

Neo-Synephrine®. See Phenylephrine hydrochloride later in this entry.

Nicotine, beta-pyridyl-alpha-N-methylpyrrolidine, $C_5H_4NC_4H_7NCH_3$, is a thick, water-white levorotatory oil that turns brown upon exposure to air. The compound is hygroscopic, soluble in alcohol, chloroform, ether, kerosine, water, and oils, bp 247°C, at which point it decomposes. Specific gravity is 1.00924. Nicotine is combustible with an autoignition temperature of 243°C. Nicotine is derived by distilling tobacco with milk of lime and extracting with ether. Nicotine is used in medicine, as an insecticide, and as a tanning agent. Nicotine is commercially available as the dihydrochloride, salicylate, sulfate, and bitartrate. Nicotinic acid (pyridine-3-carboxylic acid) is a vitamin in the B complex. See also **Vitamin.**

Phenylephrine Hydrochloride. *l*-1-(meta-hydroxyphenyl-2-methylaminoethanol hydrochloride, $HOC_6H_4CH(OH)CH_2NHCN_3 \cdot HCl$, white or nearly white crystalline substance, odorless, bitter taste. Solutions are acid to litmus paper, freely soluble in water and in alcohol, mp 140–145°C. Levorotatory in solution. Phenylephrine hydrochloride is used medically as a vasoconstrictor and pressor drug. It is chemically related to epinephrine and ephedrine. Actions are usually longer lasting than the latter two drugs. The action of phenylephrine hydrochloride contrasts sharply with epinephrine and ephedrine, in that its action on the heart is to slow the rate and to increase the stroke output, inducing no disturbance in the rhythm of the pulse. In therapeutic doses, it produces little if any stimulation of either the spinal cord or cerebrum. The drug is intended for the maintenance of an adequate level of blood pressure during spinal and inhalation anesthesia and for the treatment of vascular failure in shock, shocklike states, and drug-induced hypotension, or hypersensitivity. It is also used to overcome paroxysmal supraventricular tachycardia, to prolong spinal anesthesia, and as a vasoconstrictor in regional analgesia. Caution is required in the administration of phenylephrine hydrochloride to elderly persons, or in patients with hyperthyroidism, bradycardia, partial heart block, myocardial disease, or severe arteriosclerosis. The brand name *Neo-Synephrine*® is also used to designate another product (nose drops) which does not contain phenylephrine hydrochloride. The nose drops contain xylometazoline hydrochloride.

Quinine, $C_{20}H_{24}N_2O_2 \cdot H_2O$, a bulky, white, amorphous powder or crystalline substance, with very bitter taste. It is odorless and levorotatory. Soluble in alcohol, ether, chloroform, carbon disulfide, oils, glycerol, and acids; very slightly soluble in water. Quinine is derived from finely ground cinchona bark mixed with lime. This mixture is extracted with hot, high-boiling paraffin oil. The solution is filtered, shaken with dilute sulfuric acid and then neutralized while hot with sodium carbonate. Upon cooling, quinine sulfate crystallizes out. Pure quinine is obtained by treating the sulfate with ammonia. In addition to medical uses, quinine and its salts are used in soft drinks and other beverages.

Quinine derivatives are used in therapy for mytonic dystrophy (usually weakness and wasting of facial muscles); in the treatment of certain renal (kidney) diseases. Quinine and derivatives are best known for their use in connection with malaria. Acute attacks of malaria are usually treated with oral chloroquine phosphate. The drug is given intramuscularly to patients who cannot tolerate oral medication. Combined therapy is indicated for treating *P. falciparum* infections, using quinine sulfate and pyrimethamine. A weekly oral dose of chloroquinone phosphate is frequently prescribed for persons who travel in malarious regions. The drug is taken one week prior to travel into such areas and continued for six weeks after leaving the region. Chloroquine phosphate has not proved fully satisfactory in the treatment of babesiosis, a malarialike illness caused by a parasite.

Strychnine, $C_{21}H_{24}ON_2$, hard, white crystals or powder of a bitter taste. Soluble in chloroform, slightly soluble in alcohol and benzene, slightly soluble in water and ether, mp 268–290°C, bp 270°C (5 millimeters pressure). Strychnine is obtained by extraction of the seeds of *Nux vomica* with acetic acid, followed by filtration, precipitation by an alkali, followed by final filtration. The compound is highly toxic by ingestion and inhalation. The phosphate finds limited medical use. Strychnine is also used in rodent poisons. Strychnine acts as a powerful stimulant to the central nervous system. At one time, strychnine was used in a very carefully controlled way in the treatment of some cardiac disorders. Acute strychnine poisoning resembles fully developed generalized tetanus.

Additional Reading

Arif, A., and J. Westermeyer: "Methadone Maintenance in the Management of Opioid Dependence: An International Review," Praeger, New York, 1990.

Barnes, D. M.: "The Biological Tangle of Drug Addiction," *Science*, 415 (July 22, 1988).

Bonica, J. J., and J. D. Loeser, Editors: "The Management of Pain," Lea and Febiger, Philadelphia, Pennsylvania, 1990.

Gawin, F. H.: "Cocaine Addiction: Psychology and Neurophysiology," *Science*, 1580 (March 29, 1991).

Gerstein, D. R., and L. S. Lewin: "Treating Drug Problems," *New Eng. J. Med.*, 844 (September 20, 1990).

Ghodse, H.: "Drugs and Addictive Behavior: A Guide to Treatment," Blackwell Scientific, Boston, Massachusetts, 1989.

Gillin, J. C.: "The Long and the Short of Sleeping Pills," *New Eng. J. Med.*, 1735 (June 13, 1991).

Gilpin, R. K., and L. A. Pachla: "Pharmaceuticals and Related Drugs," *Analytical Chemistry*, 130R (June 15, 1991).

Grobbee, D. E., et al.: "Coffee, Caffeine, and Cardiovascular Disease in Men," *New Eng. J. Med.*, 1026 (October 11, 1990).

Holden, C.: "Past and Present Cocaine Epidemics," *Science*, 1377 (December 15, 1989).

Holloway, M.: "R_x for Addiction," *Sci. Amer.*, 94 (March 1991).

Masto, D. F.: "Opium and Marijuana in American History," *Sci. Amer.*, 40 (July 1991).

Nahas, G. G., and H. Peters: "Cocaine: The Great White Plague," Paul S. Eriksson, Middlebury, Vermont, 1989.

Oates, J. A., and A. J. J. Wood: "Drug Therapy," *New Eng. J. Med.*, 1017 (October 3, 1991).

Tonnesen, P., et al.: "A Double-Blind Trial of a 16-Hour Transdermal Nicotine Patch in Smoking Cessation," *New Eng. J. Med.*, 311 (August 1, 1991).

ALKALOSIS. A condition of excess alkalinity (or depletion of acid) in the body, in which the acid-base balance of the body is upset. The hydrogen ion concentration of the blood drops below the normal level, increasing the pH value of the blood above the normal 7.4. The condition can result from the ingestion or formation in the body of an excess of alkali, or of loss of acid. Common causes of alkalosis include: (1) overbreathing (hyperventilation) where a person may breathe too deeply for too long a period, consequently washing out carbon dioxide from the blood, (2) ingestion of excessive alkali, as for example an over-dosage of sodium bicarbonate possibly taken for the relief of gastric distress, and (3) excessive vomiting, which leads to loss of chloride and retention of sodium ions. The usual, mild symptoms of alkalosis are restlessness, possible numbness or tingling of the extremities (hands and feet), and generally increased muscular irritability. Only in extreme cases, tetany (muscle spasm) and convulsions may be evidenced.

See also **Acid-Base Regulation (Blood); Blood; Kidney and Urinary Tract;** and **Potassium and Sodium (In Biological Systems).**

ALKANE. One of the group of hydrocarbons of the paraffin series, e.g., methane, ethane, and propane. See also **Organic Chemistry.**

ALKENE. One of a group of hydrocarbons having one double bond and the type formula C_nH_{2n}, e.g., ethylene and propylene. See also **Organic Chemistry.**

ALKYD RESINS. The esterification of a polybasic acid with a polyhydric alcohol yields a thermosetting hydroxycarboxylic resin, commonly referred to as an alkyd resin. Some common uses of alkyds are military switchgear, electrical terminal strips, electrical relay housings and bases, and television tuner segments. The resin frequently is combined with organic and inorganic fillers. These impart desired electrical and physical properties and advantageously influence the molding characteristics. Among the advantages of alkyds are rapid curing, with no volatiles emitted during the cure cycle; low molding pressures; and high production rates on compression or transfer presses and in injection molding machines. The resins have very good dimensional stability and electrical properties.

The mineral-filled grades, sometimes modified with cellulose for reducing specific gravity and cost, are used in small switch housings, automotive ignition parts, and electronic component bases.

Alkyd resins are furnished in three major forms: (1) *fibrous*, in which the resins are compounded with long glass fibers (about 1/2-inch; 12 millimeters) and have medium strength; (2) *rope*, which is a medium-impact material and conveniently handled and processed; and (3) *granular*, in which the resins are compounded with other fibers, such as glass, asbestos, and cellulose (length about 1/16-inch; 2 millimeters). A commonly used member of the alkyd resin family is made from phthalic anhydride and glycerol. These resins are hard and possess very good stability. Where maleic acid is used as a starting ingredient, the resin has a higher melting point. Use of azelaic acid produces a softer and less brittle resin. Very tough and stable alkyds result from the use of adipic and other long-chain dibasic acids. Pentaerythritol may be substituted for glycerol as a starting ingredient.

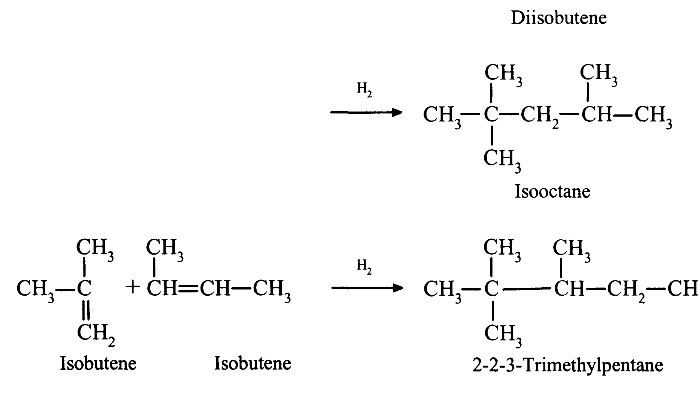
Alkyd resins are extensively used in paints and coatings. Some advantages include good gloss retention and fast drying characteristics. However, most unmodified alkyds have low chemical and alkali resistance. Modification with esterified rosin and phenolic resins improves hardness and chemical resistance. Styrene and vinyl toluene improve hardness and toughness. For high-temperature coatings (up to about 450°F; 232°C), copolymers of silicones and alkyds are used. Such coatings include stove and heating equipment finishes. To obtain a good initial gloss, improved adhesion, and exterior durability, acrylic monomers can be copolymerized with oils to modify alkyd resins. Aromatic acids, such as benzoic or butylbenzoic, also have been used with alkyd resins in coatings.

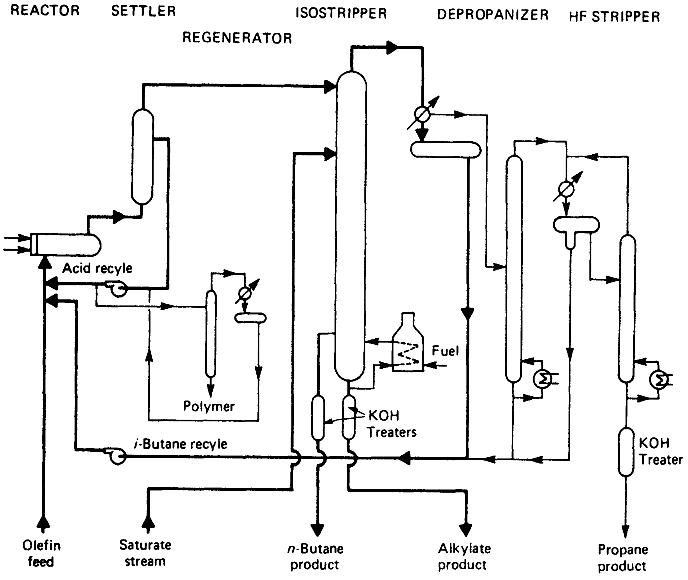
ALKYL. A generic name for any organic group or radical formed from a hydrocarbon by elimination of one atom of hydrogen and so producing a univalent unit. The term is usually restricted to those radicals derived from the aliphatic hydrocarbons, those owing their origin to the aromatic compounds being termed "aryl."

ALKYLATING AGENTS. See **Cancer and Oncology.**

ALKYLATION. Addition of an alkyl group. These reactions are important throughout synthetic organic chemistry; for example, in the production of gasolines with high antiknock ratings for automobiles or for use in aircraft.

The nature of the products of these reactions, as well as the yields, depend upon the catalysts and physical conditions. The reactions below have been written to show two combination reactions of two isobutene molecules, one yielding diisobutene, which reduces to isooctane, and the other yielding a trimethylpentane by a direct reduction reaction.





Hydrofluoric acid alkylation unit. (*UOP Process Division.*)

Specifically, the term is applied to various methods, including both thermal and catalytic processes, for bringing about the union of paraffin hydrocarbons with olefins. The process is especially effective in yielding gasolines of high octane number and low boiling range (aviation fuels).

In the petroleum industry, catalytic cracking units provide the major source of olefinic fuels for alkylation. A feedstock from a catalytic cracking units is typified by a C_3/C_4 charge with an approximate composition of: propane, 12.7%; propylene, 23.6%; isobutane, 25.0%; *n*-butane, 6.9%; isobutylene, 8.8%; 1-butylene, 6.9%; and 2-butylene, 16.1%. The butylenes will produce alkylates with octane numbers approximately three units higher than those from propylene.

One possible arrangement for a hydrofluoric acid alkylation unit is shown schematically in the accompanying figure. Feedstocks are pretreated, mainly to remove sulfur compounds. The hydrocarbons and acid are intimately contacted in the reactor to form an emulsion, within which the reaction occurs. The reaction is exothermic and temperature must be controlled by cooling water. After reaction, the emulsion is allowed to separate in a settler, the hydrocarbon phase rising to the top. The acid phase is recycled. Hydrocarbons from the settler pass to a fractionator which produces an overhead stream rich in isobutane. The isobutane is recycled to the reactor. The alkylate is the bottom product of the fractionater (isostripper). If the olefin feed contains propylene and propane, some of the isostripper overhead goes to a depropanizer where propane is separated as an overhead product. A hydrofluoric acid (HF) stripper is required to recover the acid so that it may be recycled to the reactor. HF alkylation is conducted at temperature in the range of 24–38°C (75–100°F).

Sulfuric acid alkylation also is used. In addition to the type of acid catalyst used, the processes differ in the way of producing the emulsion, increasing the interfacial surface for the reaction. There also are important differences in the manner in which the heat of reaction is removed. Often, a refrigerated cascade reactor is used. In other designs, a portion of the reactor effluent is vaporized by pressure reduction to provide cooling for the reactor.

ALKYNES. A series of unsaturated hydrocarbons having the general formula C_nH_{2n-2}, and containing a triple bond between two carbon atoms. The simplest compound of this series is acetylene $HC{:}CH$. Formerly, the series was named after this compound, namely the *acetylene* series. The latter term remains in popular usage. Particularly, the older names of specific compounds, such as acetylene, allylene $CH_2C{:}CH$, and crotonylene $CH_3C{:}CCH_3$, persist. These compounds also are sometimes called *acetylenic hydrocarbons*. In the alkyne system of naming, the "yl" termination of the alcohol radical corresponding to the carbon content of the alkyne is changed to "yne." Thus, C_2H_2 (acetylene by the former system) becomes *ethyne* (the "eth" from ethyl(C_2)); and C_4H_6 (crotonylene by the former system) becomes *butyne* (the "but" from butyl(C_4)). See also **Organic Chemistry.**

ALLANITE. Allanite is a rather rare monoclinic mineral of somewhat variable but quite complex chemical composition, perhaps represented satisfactorily by the formula $(Ce, Ca, Y)_2(Al, Fe)_3Si_3O_{12}\ (OH)$. The color of the fresh mineral is black but it is usually brown or yellow with a coating of some alteration product; often the altered crystals have the appearance of small rusty nails. It occurs characteristically in plutonic rocks like granite, syenite or diorite and is found in large masses in pegmatites. Localities in the United States are Essex and Orange Counties, New York, Franklin, New Jersey, Amherst County, Virginia, and Llano County, Texas. The slender prismatic crystals are sometimes called orthite. Allanite was named for its discoverer, T. Allan. Orthite was so named from the Greek word meaning straight, in reference to the straight prisms, a common habit of this mineral.

ALLANTOIS. A sac-like outgrowth of the hind gut of the embryo found only in reptiles, birds and mammals. In reptiles and birds it serves as a respiratory organ and receives waste matter, and in mammals it forms part of the placenta through which all interchange with the blood of the mother during embryonic development is carried out.

ALLEGHENY OROGENY. The term for an event which caused deformation of the rocks of the Valley and Ridge province, and those of the adjacent Allegheny Plateau in the central and southern Appalachians. It is believed that most of the orogeny occurred late in the Paleozoic, but some phases may have extended into the early Triassic. Use of this term is preferred to the more inclusive term, Appalachian Revolution.

ALLELE. Also termed *allelomorph*, one of two or more forms of a gene that occupies a particular locus on a chromosome. In humans there is a gene at a particular locus on a chromosome which produces an enzyme essential to the breakdown of the amino acid phenylalanine. Since there are two of each kind of chromosome in a cell, a person will normally have two genes at this locus, one on each of the two chromosomes. If at least one of these genes is the allele for enzyme production, the person will be normal. If both genes are the allele for lack of enzyme production, the phenylalanine will not be broken down and the person will suffer from phenylketonuria (PKU). Hence, we say that the gene for enzyme production is the dominant allele and the gene for no enzyme is the recessive allele.

In many cases more than two alleles are known for a particular locus on a chromosome. These are called *multiple alleles.* One locus on the X-chromosome of the fruit fly, *Drosophila*, includes a number of alleles which can cause the eye to range in color from white to a deep red. The gene for red (wild type) eye is dominant over the other alleles, but there is some intermediate inheritance when two of the other alleles are present. An individual normally will carry no more than two alleles of a series.

A.C. Vickery, Ph.D., Associate Professor, College of Public Health, University of South Florida, Tampa, Florida.

ALLELOPATHIC SUBSTANCE. A material contained within a plant that tends to suppress the growth of other plant species. The alkaloids present in several seed-bearing plants are believed to play an allelopathic role. Other suspected allelopathic substances contained in some plants include phenolic acids, flavonoids, terpenoid substances, steroids, and organic cyanides.

ALLERGY. A reaction to a specific substance in an individual who is sensitive to that substance. If the material is borne by the wind and produces symptoms of allergy when it comes into contact with the mucous membranes of the eyes and respiratory tract, the victim is said to have *hay fever*. There are two main types: (1) *seasonal hay fever* is the most common and occurs during the spring and summer seasons as the result of pollen from various trees, grasses, and weeds (the most common cause); and (2) *nonseasonal or perennial hay fever* that may be caused by allergic reactions to house pets, foods, dust, and numerous other substances. In most cases, the causative substance is inhaled.

It is believed that, in part, the reason some people react to various allergens while others do not is due to hereditary factors. A primary factor is thought to be the predisposition of allergic individuals to produce inappropriately high levels of *IgE* to potential allergens. Hay fever is not inherited, but the tendency to develop an allergic disease may be inherited. Other contributing factors include psychic stress, infections, and endocrine disturbances. Any of these factors may trigger an attack. It should be stressed that the term *hay fever* is a misnomer, in that the condition is not ordinarily associated with hay or a fever. The term was used by an English physician (Bostock), himself a victim, in a report in 1812. His symptoms occurred during the haying season.

Nearly all cases of seasonal hay fever are caused by pollen. The pollen from many sweet-scented flowers is disseminated by insects, but that causing hay fever is usually spread by the wind. Since many plants produce large quantities of pollen, a sensitive person can easily become overexposed and experience an immediate reaction. Seasonal hay fever can be caused by three different groups of plants, each of which has a somewhat different season. Trees produce pollen that causes hay fever during the months of April and May in the Northern Hemisphere (October and November in the Southern Hemisphere). Various grasses are responsible for much of the hay fever that occurs during the first half of May to the first part of July (first half of November to first part of January in Southern Hemisphere). From the middle of August to October (middle of February to April in Southern Hemisphere), weeds are active pollen producers.

Usually, the least severe and least common form of seasonal hay fever is that which is induced by pollen from trees. It is estimated that 10% of all seasonal hay fever is caused by tree pollen. The oak tree is the most common tree causing hay fever and is found widely in the United States. Three other offenders, the cottonwood, the cedar, and the poplar, grow profusely in the southern and southwestern regions of the United States. Other trees with allergy-producing pollen include the birch, alder, hickory, black walnut, beech, maple, hackberry, sycamore, mulberry, and elm.

Grasses appear to account for about 35% of all cases of seasonal hay fever. The three most common grasses involved are timothy, Bermuda, and June (blue grass). There are eighty varieties of grasses found in various sections of the United States that have a common antigen. Therefore, a person sensitive to one type of grass is usually sensitive to all types of grasses.

Weeds are prolific pollen producers, some varieties producing more than 100,000 pollen grains from a single plant. The most causative agents include various species of ragweed and the thistle. Other weeds involved, but not on a wide scale, include goosefoot, buckwheat, marsh elder, rabbit bush, cocklebur, hemp, and pigweed.

Nonseasonal hay fever may be caused by the hair, feathers, or dander of a household or farm pet. Some individuals are sensitive to feathers in pillows or to kapok, the fibers used as filling for mattresses. House dust, especially in the bedroom, is a common causative agent. Workers in mills where wheat or corn is ground often inhale the flourlike powder, which produces irritation of the nasal mucous membranes. Reproductive cells of some fungi, when inhaled, may stimulate an allergic response.

Nonseasonal hay fever may be caused by certain foods, in particular eggs, chocolate, milk, coffee, or shellfish. Aspirin and quinine may be causative agents. Nonseasonal hay fever may be continuous or spasmodic, depending upon the length of contact with the exciting factor. The symptoms of nonseasonal hay fever tend to be less severe than those of the seasonal variety. At their mildest, they may consist of slight nasal congestion with sniffing, a tendency to an itchy nose, postnasal drip, and mouth breathing.

While the symptoms of hay fever are not difficult to recognize, the actual type of hay fever is not so easily determined. The prior history of a patient is of prime importance. Were there allergic reactions during childhood? Family history is helpful to determine any disposition toward the disease. In particular, details concerning the patient's living habits and possible exposure to various causative agents are important.

Skin tests (also known as "scratch" tests or immediate hypersensitivity tests) are frequently made to obtain needed information. A series of small intradermal injections of suspected allergens are made in the skin and the area is examined 30 minutes later. If the test is positive, a large, red, itching wheal will appear. If no reaction occurs after half an hour, the material is removed and the reaction is regarded as negative. The tests are not infallible. Some individuals have more sensitive skin than others. Consequently, false positive reactions are not uncommon. A second type of test, used less frequently, is an eye test. A drop of solution of pollen extract is dropped into the eye. If the reaction is positive, a condition similar to the eye symptoms of hay fever results.

In lieu of living in areas where the pollen count is low (often difficult for a variety of employment, economic, family and other reasons), air conditioning of the home or office is helpful. A variety of drugs provide relief of hay fever symptoms. Antihistamines are often used and help a high percentage of individuals. Ephedrine-like drugs, taken at night and either alone or in combination are effective in lessening early morning symptoms. Corticosteroids are useful in severe cases where other agents are not effective.

Preventive treatment also is available. It is designed to acclimate a person's body to the irritant, once identified. The treatment is commenced approximately three months prior to the usual onset of an attack, known from past history of the patient. The treatment consists of the injection of the pollen extract at intervals of about a week. The dosage is gradually increased until the largest dose is given at about the time the symptoms usually start. From then on, the dosage remains the same until the end of the season. At the beginning of treatment, there may be a mild reaction on the spot where the injection was given. Subsequent dosage is regulated by the severity of this reaction. The pollen extract treatment has been beneficial in many cases.

The treatment of patients with nonseasonal hay fever consists of completely avoiding the substance or substances which cause the attack. Should contact be absolutely necessary, the physician may prescribe the allergen extract treatment, particularly if the patient's symptoms are so severe as to incapacitate him in his work.

In some allergic conditions there may be upsets of the digestive system, and occasionally severe headaches may result. In all of these cases, there may be some accompanying change in the skin, but in other commonly encountered allergies, the skin changes constitute about the only symptoms. The specific skin symptom may bear very little relation to the cause of the allergy; a particular antigen produces varying types of responses in the skin of different individuals.

One of the most common skin changes associated with allergy is simply a reddening (*erythema* or *hypermia*), caused by increased amounts of blood in the lower layers of the skin due to localized capillary dilatation. Reddened areas of this type may be restricted to a small area of the body, or may be general over its surface. They turn white when subjected to pressure from a finger, seldom are long-lasting, and either disappear within a few days, or progress into some other type of symptom.

Hives (*urticaria*) is a common skin condition in which whitish or reddish, slightly elevated areas of the skin appear. These wheals may be small, like pimples (*papules*), or much larger patches or streaks (*welts*). They generally cover the entire body, being most common on areas covered by clothing. Hives are caused by the accumulation of tissue fluids (*edema*) beneath the epidermis in areas seen as wheals. The condition generally arises rapidly, may last for an hour or so, and then disappears as quickly as it came, if its cause has been removed. See also **Urticaria.**

Another symptom frequently associated with allergy is known as *eczema*. There is a reddening of the skin, followed by the appearance of minute blisters or *vesicles*. These vesicles become larger and are generally accompanied by intense itching. In acute cases, these blisters break and exude a fluid which forms a crust on the skin. The crust then flakes off, frequently as the result of a secondary inflammation of the skin. Eczema may cover any area of the body, and is one of the most severe of all allergic symptoms. Eczema-type reactions of the skin may result also from some infections and as the result of various nervous conditions.

Other symptoms occasionally seen as the result of an allergy include *nodules*, which are small hard bodies beneath the skin, and large blisters (*blebs* or *bullae*). As the result of the various skin changes which occur, secondary lesions eventually may develop. These include abrasions or erosions, fissures or cracks, ulcers, and scars. These secondary lesions are seldom encountered when the patient receives prompt treatment and the cause of the allergy is determined and removed.

Food allergies in infants frequently result in a severe eczema, and are most often caused by egg white, milk, wheat, oats, barley, and corn. Since eczema may result the first time an infant eats egg white or some other of these foods, it seems possible that sensitization of a child may have occurred while it was receiving its nourishment through the placenta before birth. Infantile eczema most often appears in the second or third month of life and may disappear spontaneously by the end of the second year, with no remaining signs of the food hypersensitivity. The condition may appear again or become worse following vaccination, colds, or eruption of the teeth. Sensitivity to egg, wheat, and milk usually occurs less frequently with increasing age, and disappears almost completely between the fourth and twelfth years.

A large number of chemical substances when taken into the body or applied to the body's surface are capable of producing severe allergic symptoms. Not only are such skin conditions encountered as the result of some medicine to which the body has become sensitized, but they also occur as the result of contact with various industrial chemicals.

Skin eruptions caused by drugs differ somewhat from other allergies, in that they frequently manifest brighter colors, appear suddenly, occur symmetrically on the body, are frequently extensive, and do not generally produce any other body disturbances. Most symptoms disappear after administration of the drug is stopped. Skin eruptions caused by iodides and bromides disappear more slowly and those caused by arsenic hypersensitivity may appear long after the drug has been taken and may last indefinitely. Hypersensitivity to phenolphthalein (used in some laxatives) also may produce an inflammation which lasts long after administration of the drug has been stopped.

Among the more common drugs that may cause eruptions might be listed acetanilide, amidopyrine, antipyrine, arsenic compounds, aspirin, atabrine, barbituric acid derivatives, benzoic acid, benzocaine, opium and morphine, penicillin, phenobarbital, phenolphthalein, quinine, salicylic acid, sulfonamides, and turpentine. Except for reactions to penicillin, it is evident that, when one considers the number of persons to whom they are administered without ill effects, allergy to any one of these drugs is a relatively rare condition. Included among the various medicinal preparations which are capable of producing allergies should be mentioned the various sera and other animal products. When various immunizing sera, such as tetanus antitoxin, are repeatedly injected into an individual, they occasionally produce a sensitive condition as the result of the development of antibodies against the proteins in the serum. In some acute cases, the entire body may react violently to a further administration of the same serum. The dangerous condition which occurs within a few moments in such cases is known as *anaphylactic shock*. Modern methods of preparing the sera for injection have caused a marked decrease in the incidence of this condition.

Workers sometimes develop a hypersensitivity to materials to which they are constantly exposed, as bakers to flour, barbers to quinine (hair tonics), dentists to Novocaine, painters to linseed oil, and so on. Various soaps and detergents are also common allergens, although these agents are more often responsible for *primary irritant dermatitis*, a condition easily confused with true allergy. Toilet preparations, cosmetics, clothing, and a host of other substances can set up an allergic reaction in some people. Insect bite hypersensitivity is common. In some individuals, a simple mosquito bite may produce a large and painful swelling out of all proportion to that seen in most other persons. Bites or stings by bees, wasps, bedbugs, lice, fleas, gnats, caterpillars, and various marine fishes and other animals may produce extreme reactions in some few individuals who have previously been sensitized to the allergenic materials of the particular species.

Heat, cold, and light may be the direct cause of burn, chapping, and sunburn, but in some sensitive persons, they may produce allergic skin changes. These usually take the form of hives. In most cases, the symptoms subside rapidly after the cause has been removed.

Mental and emotionally induced allergic symptoms also may appear. The mechanism is not well understood, but it is believed that strong emotions may release various chemical substances into the bloodstream, substances which are capable of sensitizing the body and which act as allergens.

See also **Immune System and Immunology.**

ALLIGATION. A simple mathematical method for calculating the correct proportioning of the ingredients of a mixture or, in general, the value of a property of a mixture from the values of that property in its components. It is based upon the formula

$$P_{xy} = \frac{xX + yY}{x + y}$$

in which P_{xy} is the value of a property of the mixture, X and Y are its values in the components, and x and y are the proportions of the components. It assumes, of course, no change in properties on mixing.

ALLIGATOR. See **Crocodiles and Alligators.**

ALLIUM (*Liliaceae*). A large genus whose species are found widely. Some 75 species are found in North America, especially in the western states. All are bulbous plants with flat or tubular leaves, and with spherical heads or umbels of variously colored flowers. Particularly important cultivated species are the onion, *Allium cepa*; leek, *Allium porrum*; garlic, *Allium sativum*; and chives, *Allium schoenoprasum*. One European species now extensively introduced in the United States is the common weed, field garlic, *Allium vineale*, which (if eaten by cows) noticeably flavors milk and butter.

ALLOBAR. A form of an element differing in atomic weight from the naturally occurring form, hence a form of element differing in isotopic composition from the naturally occurring form.

ALLOCHROMATIC. With reference to a mineral that, in its purest state, is colorless, but that may have color due to submicroscopic inclusions, or to the presence of a closely related element that has become part of the chemical structure of the mineral. With reference to a crystal that may have photoelectric properties due to microscopic particles occurring in the crystal, either present naturally, or as the result of radiation.

ALLOCHROMY. Any fluorescence, or reradiation of light, in which the wavelength (and hence color) of the emitted light differs from that of the absorbed light.

ALLOCHTHONOUS. A term proposed by Gümbel in 1888 for sedimentary rocks whose constituents have been transported and deposited at some distance from their place of origin. The bulk of the sedimentary rocks are of this type. The term is now used most commonly in reference to masses of rock transported considerable distances by tectonic movements. See **Overthrust.**

ALLOGENIC (Ecology). Successive ecologic events or conditions that result from factors which arise from outside the natural community and thus alter the more localized situation. The term may be applied, for example, to an allogenic drought of very long duration.

ALLOGENIC (Geology). Minerals and rock constituents derived from pre-existing rocks that have been transported from their original site. The term also applies to a stream (allogenic stream) which is fed by water from a distant terrain. An example would be a stream that originates in a humid or glacial region and that later flows through an arid or desert region.

ALLOGYRIC BIREFRINGENCE. A beam of plane-polarized light may be regarded as the resultant of two equal beams of circularly polarized light, one right-handed and the other left-handed. The phenomenon of optical rotation may be represented by assuming that in optically active media, circularly polarized light is transmitted unchanged, but the velocity of left-handed circularly polarized light is not the same as that of right-handed circularly polarized light. Fresnel demonstrated this difference directly, and the phenomenon is called allogyric birefringence.

ALLOMERISM. A property of substances that differ in chemical composition but have the same crystalline form.

ALLOMORPHISM. A property of substances that differ in crystalline form but have the same chemical composition.

ALLOTYPE. An animal or plant fossil selected, as a species or subspecies, as illustrating morphological details not shown in the holotype.

ALLOYS. Traditionally, an alloy has been defined as a substance having metallic properties and being composed of two or more chemical elements of which at least one is a metal (ASM). Although this still covers the general use of the word, in recent years alloy also has been used in connection with other, non-metallic, materials. Most metals are soluble in one another in their liquid state. Thus alloying procedures usually involve melting. However, alloying by treatment in the solid state without melting can be accomplished in some instances by such methods as powder metallurgy. When molten alloys solidify, they may remain soluble in one another, or may separate into intimate mechanical mixtures of the pure constituent metals. More often, there is partial solubility in the solid state and the structure consists of a mixture of the saturated solid solutions. Another important type of solid phase is the intermetallic compound which is characterized by hardness and brittleness and usually has only limited solid solubility with the other phases present. The interactions of two or more elements both in the liquid and solid state are effectively characterized by phase diagrams. Where only two principal materials are involved, *binary alloy* is the term used. Three principal ingredients are referred to as *ternary alloy.* Beyond three components, the material may be referred to as a multicomposition system or alloy.

The decade of the 1980s has witnessed the development of hundreds of new alloys, involving not only the traditional metals, but much greater use of the less common chemical elements, such as indium, hafnium, etc. In this encyclopedia, alloys of a chemical element are discussed mainly under that particular element, or in an entry immediately following. Also check alphabetical index.

In addition to the appearance of numerous new alloys, sometimes called superalloys, recent developments in this field include many relatively new processes and methodologies, such as electron beam refining, rapid solidification, single-crystal superalloys, and metallic glasses, among others. Some of these are described in separate articles in this encyclopedia.

Motivation for the development of new alloys is found in nearly all consuming areas, but particular emphasis has been given to the expanding and increasingly demanding requirements of the aircraft and aerospace industries, including much attention directed to the lighter elements (titanium, aluminum, etc.); the needs of the military; the very difficult requirements of jet engine parts; and the electronics industry where much attention has been directed toward the less common metals. Within recent years, metallurgists also have come to appreciate that processing of alloys can be of as much importance as the elements which they contain. New processes have been developed during the past decade or so, including rapid solidification, electron beam refining, and many others. Metal alloy research also has been impacted by the rapidly and continuously expanding science and art of making composites, often involving ceramics, graphite, organics, etc. in addition to metals. Knowledge of alloys not only must assist the applications of simply the alloys themselves, but also how the alloys perform in a composite part.

Predicting the Performance of Alloys. It is well known that many important alloy combinations have properties that are not easy to predict, simply on the basis of knowledge of the constituent metals. For example, copper and nickel, both having good electrical conductivity, form solid-solution type alloys having very low conductivity, or high resistivity, making them useful as electrical resistance wires. In some cases very small amounts of an alloying element produce remarkable changes in properties, as in steel containing less than 1% carbon with the balance principally iron. Steels and the age-hardening alloys depend on heat treatment to develop special properties such as great strength and hardness. Other properties which can be developed to a much higher degree in alloys than in pure metals include corrosion-resistance, oxidation-resistance at elevated temperatures, abrasion- or wear-resistance, good bearing characteristics, creep strength at elevated temperatures, and impact toughness. However, solid state physics has been successful in explaining many of these properties of metal and alloys.

There are various types of alloys. Thus, the atoms of one metal may be able to replace the atoms of the other on its lattice sites, forming a substitutional alloy, or solid solution. If the sizes of the atoms, and their preferred structures, are similar, such a system may form a continuous series of solutions; otherwise, the miscibility may be limited. Solid solutions, at certain definite atomic proportions, are capable of undergoing an order-disorder transition into a state where the atoms of one metal are not distributed at random through the lattice sites of the other, but form a superlattice. Again, in certain alloy systems, intermetallic compounds may occur, with certain highly complicated lattice structures, forming distinct crystal phases. It is also possible for light, small atoms to fit into the interstitial positions in a lattice of a heavy metal, forming an interstitial compound.

In this encyclopedia alloys of chemical elements of alloying importance are discussed under that particular element.

Alloy Phase Diagram Data Programme

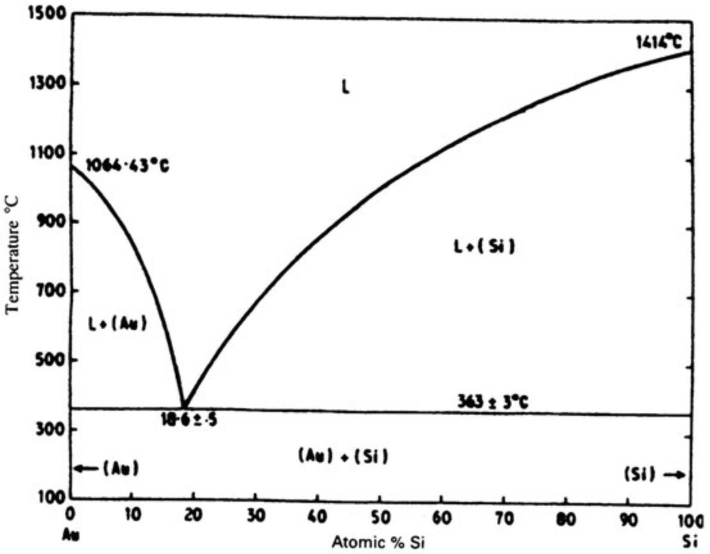
Alloy phase diagrams have been known since 1829 when the Swedish scientist Rydberg, who observed the thermal effects that occur during the cooling of binary and ternary alloys from a molten condition. Gibbs many years later published a treatise on the theory of heterogeneous equilibria. The practical importance of phase diagrams awaited the development of the phase diagram for the iron-carbon system, which became central to the metallurgy of steel. See **Iron Metals, Alloys, and Steels.** With the development over the years of scores of binary alloys, considering the number of chemical elements involved, and then followed by ternary and much more complex alloys—with many hundreds of professionals in the metal sciences contributing knowledge—the problems of collecting and of easily locating such information took on formidable proportions. The start of an effective database was the publication of a compilation, by Hansen in 1936, of information gleaned from the literature on 828 binary systems, for which sufficient data were available to construct phase diagrams for 456 binary systems. An English version of the German works, updated to some extent, appeared in 1958. This compilation included 1324 binary systems and 717 binary phase diagrams. A supplementary volume by R. P. Elliott brought the number of binary phase diagrams to 2067 and a later work by F. A. Skunk (1969) included data on 2380 systems. These efforts became key reference works for metallurgists concerned with alloy development and alloy applications. For obvious reasons, information on ternary phase diagrams and other multicomponent systems was far less satisfactory.

To improve this important metallurgical database, the National Bureau of Standards (U.S.) and the American Society for Metals, after many prior deliberations, each signed a memorandum of agreement to proceed with a data programme for alloy phase diagrams, concentrating on binary systems. As early as 1975, T. B. Massalski (Carnegie-Mellon University), the current chairman of the programme, observed that a knowledge of phase diagram data is basic for the technological application of metals and alloys; that any programme to provide critically evaluated data would have to be a worldwide enterprise because there is too much work for any institution or organization, or even any country, to accomplish the task alone; that the programme should deal with binary and multicomponent systems; that a computerized bibliographic database should be developed; that computer technology should be used to provide data for the generation of phase diagrams at remote terminals; and that funding for the programme should be sought. As of the late 1980s, many of these objectives have been achieved. Among organizations not previously mentioned, cooperation has been given by the Institute of Metals (U.K.), the U.K. Universities' Science Research Council, the Max-Planck Institut für Metallforschung (Stuttgart), among several other sources, including funding from various interested corporations.

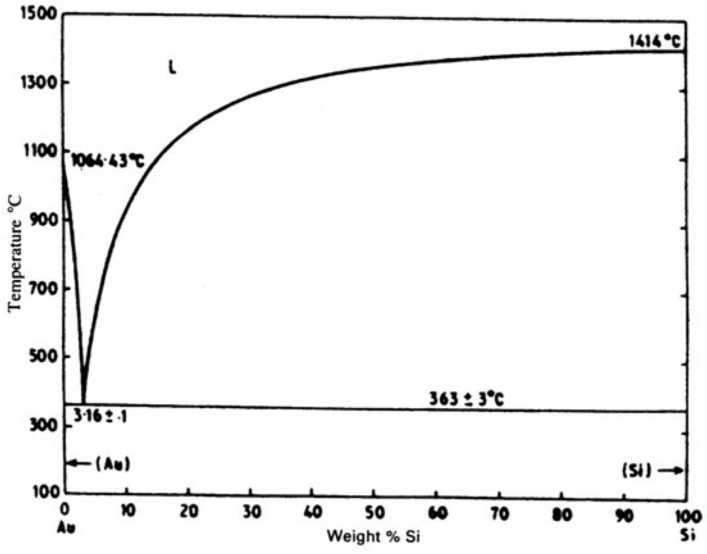
The massive task involved required that the chore be broken down for handling by category editors, of which there are approximately thirty. The first comprehensive publication to be released thus far is from the American Society for Metals (ASM International), entitled

"Binary Alloy Phase Diagrams," which contains up-to-date and comprehensive phase diagram information for more than 1850 alloy systems, representing the first major release of critically evaluated phase diagrams since 1969.

Importance of Phase Diagrams. As pointed out by Massalski and Prince (1986), phase diagrams are graphic displays of the thermodynamic relationships of one or more elements at different temperatures and pressures. It has been stated that phase diagrams are to the metallurgist what anatomy is to the medical profession or cartography to the explorer. To explain this analogy, reference is made to a specific phase diagram (Fig. 1). This gold-silicon (Au-Si) phase diagram is a two-dimensional mapping of the phases that form between Au and Si as a function of temperature and of alloy composition. In Fig. 1(a), the alloy composition is defined in terms of the percentage of atoms of Si in the alloy. It will be noted that a dramatic lowering of the freezing point of pure gold (1064.43°C) occurs upon the addition of silicon. Conversely, there is a more regular depression of the freezing point of Si (1414°C) upon the addition of Au.



(a)



(b)

Fig. 1. Representative binary alloy phase diagrams: (a) the Au-Si phase diagram with compositions in atomic percent; (b) the Au-Si phase diagram with compositions in weight percent. (*ASM News*.)

The two upper curves, termed the *liquidus curve*, define the temperatures at which Au-Si alloys begin to solidify. The curves meet at 363°C at an alloy composition containing 18.6 atomic percent Si. At this temperature, all Au-Si alloys, irrespective of composition, complete their solidification by the eutectic separation of a fine mixture of Au and Si from the liquid phase containing 18.5 atomic percent Si. The horizontal line at 363°C is called the *solidus* because below such a line all of the alloys are completely solid.

The effect of presenting the alloy composition in terms of the weight percentage of Si is shown in Fig. 1(b). The liquidus curves drop even more dramatically towards the Au-rich side of the phase diagram and the eutectic liquid at 363°C contains only 3.16 weight percent Si. The movement of the eutectic composition, the lowest-melting alloy composition, from 18.6 percent Si of Fig. 1(a) to 3.15 weight percent Si in Fig. 1(b) simply reflects the great difference in the atomic weights of Au and Si.

Massalski and Prince selected this particular phase diagram because of its simplicity for illustration and because this particular phase diagram is of considerable importance in the semiconductor device industry. Silicon chips are frequently bonded to a heat sink, using a gold or more frequently a Au-Si alloy foil placed between them. Upon heating above 363°C, the Au reacts with Si to form a brazed joint between the silicon chip and the heat sink.

Phase diagrams are condensed presentations of a large amount of information. They provide quantitative information on the phases present under given conditions of alloy composition and temperature and, to the experienced metallurgist, a guide to the distribution of the phases in the microstructure of the alloy. They also dictate what alterations in phase constitution will occur with changing conditions, whether these be alteration of alloy composition, temperature, pressure or atmosphere in equilibrium with the material.

Further, the phases present in an alloy and their morphological distribution within the microstructure, define the mechanical, chemical, electrical, and magnetic properties that may be achievable. Thus, we have the essential link between the engineering properties of an alloy and its phase diagram. Indeed, a distinctive feature of metallurgy as a profession is that it is primarily concerned with the relationship between the constitution and the properties of alloys. Phase diagram data are key elements to understanding, and thereby controlling, the properties of alloys.

Broad Categories of Alloys

Although there are thousands of alloys, with many new alloys appearing each year, there are certain traditional alloys that serve the vast majority of materials needs. The bulk of new alloys, although extremely important, are frequently application-specific. The broad classes are described briefly as follows.

CAST FERROUS METALS

Gray, Ductile, and High-Alloy Irons

In gray iron, most of the contained carbon is in the form of graphite flakes, dispersed throughout the iron. In ductile iron, the major form of contained carbon is graphite spheres which are visible as dots on a ground surface. In white iron, practically all contained carbon is combined with iron as iron carbide (cementite), a very hard material. In malleable iron, the carbon is present as graphite nodules. High-alloy irons usually contain an alloy content in excess of 3%.

Malleable Iron

The two main varieties of malleable iron are ferritic and pearlitic, the former more machinable and more ductile; the latter stronger and harder. Carbon in malleable iron ranges between 2.30 and 2.65%. Ranges of other constituents are: manganese, 0.30 to 0.40%; silicon, 1.00 to 1.50%; sulfur, 0.07 to 0.15%; and phosphorus, 0.05 to 0.12%.

Carbon and Low-Alloy Steels

Low-carbon cast steels have a carbon content less than 0.20%; medium-carbon steels, 0.20 to 0.50%; and high-carbon steels have in excess of 0.50% carbon. Ranges of other constituents are: manganese, 0.50 to 1.00%; silicon, 0.25 to 0.80%; sulfur, 0.060% maximum; and phosphorus, 0.050% maximum.

Low-alloy steels have a carbon content generally less than 0.40% and contain small amounts of other elements, depending upon the desired end-properties. Elements added include aluminum, boron, chromium, cobalt, copper, manganese, molybdenum, nickel, silicon, titanium, tungsten, and vanadium.

High-Alloy Steels

When "high-alloy" is used to describe steel castings, it generally means that the castings contain a minimum of 8% nickel and/or chromium. Commonly thought of as stainless steels, nevertheless *cast grades* should be specified by ACI (Alloy Casting Institute) designations and not by the designations that apply to similar *wrought alloys*.

WROUGHT FERROUS METALS

Carbon Steels

These steels account for over 90% of all steel production. There are numerous varieties, depending upon carbon content and method of production. In one classification, there are *killed* steels, *semikilled* steels, *rimmed* steels, and *capped* steels. These are described in considerable detail under **Iron Metals, Alloys, and Steels.**

High-Strength Low-Alloy Steels

There are several varieties, with high-yield strength depending mainly on the precipitation of martensitic structures from an austenitic field during quenching. Small additions of alloy elements, such as manganese and copper, are dissolved in a ferritic structure to obtain high strength and corrosion resistance.

Low and Medium-Alloy Steels

The two basic types are (1) *through* hardenable, and (2) *surface* hardenable. Subcategories of surface hardenable alloys include carburizing alloys, flame and induction-hardening alloys, and nitriding alloys.

Stainless Steels

A stainless steel is defined as iron-chromium alloy that contains at least 11.5% chromium. There are three major categories: (1) austenic, (2) ferritic, and (3) martensitic, depending upon the metallurgical structure. There are scores of varieties. Type 302 is the base alloy for austenitic stainless steels. Representative stainless steels in this category provide some insight as to why so many varieties are made and of how rather small changes in composition and production can bring about significant differences in the final properties of the various stainless steels. A slightly lower carbon content improves weldability and inhibits carbide formation. An increase in nickel content lowers the work hardening. By increasing both chromium and nickel, better corrosion and scaling resistance is achieved. The addition of sulfur or selenium increases machinability. The addition of silicon increases scaling resistance at high temperature. Small amounts of molybdenum improve resistance to pitting corrosion and temperature strength.

High-Temperature, High-Strength, Iron-Base Alloys

There are two general objectives in making these alloys: (1) they can be strengthened by a martensitic type of transformation, and (2) they will remain austenitic regardless of heat treatment and derive their strength from cold working or precipitation hardening. Again, there are numerous types. Considering the main types, the carbon content may range from 0.05% to 1.10%; manganese, 0.20 to 1.75%; silicon, 0.20 to 0.90%; chromium, 1.00 to 20.75%; nickel, 0 to 44.30%; cobalt, 0 to 19.50%; molybdenum, 0 to 6.00%; vanadium, 0 to 1.9%; tungsten, 0 to 6.35%; copper, 0 to 3.30%; columbium (niobium), 0 to 1.15%; tantalum, 0 to <1%; aluminum, 0 to 1.17%; and titanium, 0 to 3%.

Ultrahigh-Strength Steels

Normally a steel is considered in this category if it has a yield strength of 160,000 psi or more. The first of these steels to be produced was a chromium-molybdenum alloy steel, shortly followed by a stronger chromium-nickel-molybdenum grade.

Free-Machining Steels

Normally, the carbon content is kept under 0.10%, but as much as 0.25% carbon has little deleterious effect on machinability. Aluminum and silicon are held to a minimum (aluminum not used as a deoxidizer where machinability is extremely important). Lead, sulfur, bismuth, selenium, and tellurium (0.04%) improve machinability when in the proper combination. Sulfur improves machinability by combining with any manganese and oxygen present to form oxysulfides.

NONFERROUS METALS

Aluminum Alloys

These alloys are available as wrought or cast alloys. The principal metals alloyed with aluminum include copper, manganese, silicon, magnesium, and zinc. These alloys are discussed in considerable detail under **Aluminum Alloys.**

Copper Alloys

These alloys are available as wrought or cast alloys. The principal wrought copper alloys are the brasses, leaded brasses, phosphor bronzes, aluminum bronzes, silicon bronzes, beryllium coppers, cupro nickels, and nickel silvers. The major cast copper alloys include the red and yellow brasses, manganese, tin, aluminum, and silicon bronzes, beryllium coppers, and nickel silvers. The chemical compositions range widely. For example, a leaded brass will contain 60% copper, 36 to 40% zinc, and lead up to 4%; a beryllium copper is nearly all copper, containing 2.1% beryllium, 0.5% cobalt, or nickel, or in another formulation, 0.65% beryllium, and 2.5% cobalt.

Nickel Alloys

Although nickel is present in varying amounts in stainless steel *commercially* a high-nickel stainless steel is not categorized as a nickel *alloy*, but rather as a stainless steel. Most nickel alloys are proprietary formulations and hence designated by trade names, such as *Duranickel, Monel* (several), *Hastelloy* (several), *Waspaloy, Rene 41, Inco, Inconel* (several), and *Illium* G. The nickel content will range from about 30% to nearly 95%.

Magnesium Alloys

It is the combination of low density and good mechanical strength which provides magnesium alloys with a high strength-to-weight ratio. Again, these generally are proprietary formulations. Aluminum, manganese, thorium, zinc, zirconium, and some of the rare-earth metals are alloyed with magnesium.

Zinc Alloys

Zinc alloys are available as die-casting alloys or wrought alloys. The principal alloys used for die casting contain low percentages of magnesium, from 3.5 to 4.3% aluminum, and carefully controlled amounts of iron, lead, cadmium, and tin.

Titanium Alloys

The titanium-base alloys are considerably stronger than aluminum alloys and superior to most alloy steels in several respects. Several types are available. Alloying metals include aluminum, vanadium, tin, copper, molybdenum, and chromium.

METALLIC GLASSES

Potentially, metallic glasses and metastable crystalline alloys are the strongest, toughest, and most corrosive resistant, and the most easily magnetizable materials known to materials engineers. Metallic glasses often are quite superior to their crystalline counterparts. This is an important reason why rapid solidification technology has attracted worldwide attention. Several factors are recognized as affecting an alloy's ability to form a metallic glass. These include atomic size ratio, alloy crystallization temperature and melting point, and heat of formation of compounds. In addition to drastic supercooling of a metal alloy, metallic glasses have been made by electrodeposition or by vapor deposition. These noncrystalline metal or alloy compounds are sometimes referred to as *amorphous* alloys.

Additional Reading

Bryskin, B. D.: "Rhenium and Its Alloys," *Advanced Materials & Processes*, 22 (September 1992).

Cardonne, S. M., et al.: "Refractory Metals Forum: I: Tantalum and Its Alloys," *Advanced Materials & Processes*, 16 (September 1992).

Clement, T. P., Parsonage, T. B., and M. B. Kuxhaus: "Ti_2AlNb=Based Alloys Outperform Conventional Titanium Aluminides," *Advanced Materials & Processes*, 37 (March 1992).

Eillenauer, J. P., Nieh, T. G., and J. Wadsorth: "Tungsten and Its Alloys," *Advanced Materials & Processes*, 28 (September 1992).

Frick, J., Editor: "Woldman's Engineering Alloys," ASM International, Materials Park, Ohio, 1990.

Jackman, L. A., Maurer, G. E., and S. Widge: "New Knowledge About 'White Spots' in Superalloys," *Advanced Materials & Processes*, 18 (May 1993).

Kane, R. D., and R. G. Taraborelli: "Selecting Alloys to Resist Heat and Corrosion," *Advanced Materials & Processes*, 22 (April 1993).

Kane, R. D.: "Super Stainless Steels Resist Hostile Environments," *Advanced Materials & Processes*, 16 (July 1993).

Lai, G. Y.: "High-Temperature Corrosion of Engineering Alloys," ASM International, Materials Park, Ohio, 1990.

McCaffrey, T. J.: "Combined Strength and Toughness Characterize New Aircraft Alloy," *Materials & Processes*, 47 (September 1992).

Rioja, R. J., and R. H. Graham: "Al-Li Alloys Find Their Niche," *Advanced Materials & Processes*, 23 (June 1992).

Schweitzer, P. A.: "Corrosion Resistance Tables," 3rd Edition, ASM International, Materials Park, Ohio, 1991.

Shields, J. A., Jr.: "Refractory Metals Forum II: Molybdenum and Its Alloys," *Advanced Materials & Processes*, 28 (October 1992).

Staff: "ASM Engineered Materials Reference Book," ASM International, Materials Park, Ohio, 1988.

Staff: "Properties and Selection: Irons, Steels, and High Performance Alloys," ASM International, Materials Park, Ohio, 1990.

Staff: "Alloy Phase Diagrams," ASM International, Materials Park, Ohio, 1991.

Staff: "Properties and Selection: Nonferrous Alloys and Special-Purpose Materials," ASM International, Materials Park, Ohio, 1991.

Staff: "Advances in Aluminum and Alloys," *Advanced Materials & Processes*, 17 (January 1992).

Staff: "Product Spotlight: Chemical Analysis of Metals and Alloys," *Advanced Materials & Processes*, 31 (February 1992).

Titran, R. H.: "Refractory Metals Forum III: Niobium and Its Alloys," *Advanced Materials & Processes*, 34 (November 1992).

Voort, G. V., Editor: "Atlas of Time-Temperature Diagrams for Nonferrous Alloys," ASM International, Materials Park, Ohio, 1991.

Warner, E. A., and D. A. DeAntonio: "Controlled-Expansion Superalloy Resists Oxidation at High Temperatures," *Advanced Materials & Processes*, 51 (September 1993).

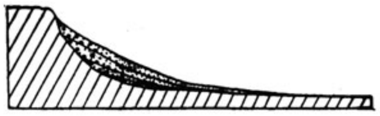
ALLSPICE. Sometimes called *pimento*, allspice is prepared from the dried, pea-size, unripened, dark reddish-brown berry of a West Indian evergreen tree (*Pimenta officinalis* L.). The tree, which reaches a height of about 20 feet (6 meters) is an evergreen and a member of the Myriaecea (myrtle) family, of which probably the best known members are bayberry trees and shrubs.

The word *pimento* is not to be confused with the word *pimentio* (see **Pepper**). The substances are not related. Also, to contribute to word confusion, allspice is sometimes called Jamaica pepper, a term which tends to associate it with the more familiar black or white pepper of the family *Piperiaceae*. There is no such association.

Allspice is well named because its essence and taste resemble that of a mixture of cinnamon, nutmeg, and cloves. Pimento oil is a fragrant essential oil distilled from allspice berries. It contains eugenol and cineol which have a carnationlike essence.

Allspice finds numerous uses in food products, including:

Bakery products—special breads and rolls; muffins; coffee cakes; spice cakes; fruit cakes; fruit, chocolate, and custard cream pies.
Beverages—cordials and liqueurs.
Condiments—catsup, chili sauce.
Confections—licorice.
Ethnic dishes—German foods.
Fruits and fruit-based products—apples; applesauce; apple drink; apricots; cranberry drink; cranberry sauce; peaches; pears; plums; preserves; spiced fruits; stewed fruits.
Meats and meat dishes—bologna; beef; frankfurters; hamburger; head cheese; meatloaf; mincemeat; pork; sausage.
Pickles.
Poultry.
Sauces—tomato sauce.
Soups—beef; consommé; tomato; vegetable.
Vegetables—beans, beets; cole slaw; spinach; squash; sweet potatoes; tomatoes; turnips.



Cross section of an alluvial fan or subaerial delta.

ALLUVIAL FAN. Also called *subaerial delta*, a cone-shaped to to delta-shaped collection of coarsely graded sediments deposited by intermittent streams that debouch from steep valleys onto a relatively gentle slope or plain. See accompanying figure. Alluvial fans may extend for many miles. Confluent fans may eventually cover and fill relatively large intermontane basins. An alluvial-fan shoreline is one in which an alluvial fan is built out into a lake or sea.

ALLUVIUM. A general term used to designate the sand, silt, and mud deposited by a stream, along its bank or upon its floodplain, during periods of high water. The word is derived from the Latin *ad*, to; and *luo*, wash. When alluvium is relatively fine-textured and contains sufficient organic matter it forms soil. Some of the oldest and richest agricultural regions are the great delta areas, such as the Nile and Euphrates.

ALLYL ESTER RESINS. The allyl radical ($CH_2CH{=}CH_2$) is the basis of the allyl family of resins. Allyl esters are based on monobasic and dibasic acids and are available as low-viscosity monomers and thermoplastic prepolymers. They are used as crosslinking agents for unsaturated polyester resins and in the preparation of reinforced thermoset molding compounds and high-performance transparent articles. All modern thermoset techniques may be used for processing allyl resins.

The most widely used allyls are the monomers and prepolymers of diallyl phthalate and diallyl isophthalate. These are readily converted into thermoset molding compounds and into preimpregnated glass cloths and papers.

Diethyleneglycol-*bis*-(allylcarbonate), marketed as *CR-39*™, is finding increasing use where optical transparency is required. It is the primary material used in the manufacture of *plastic lenses for eyewear* because of its light weight, dimensional stability, abrasion resistance, and dye-ability. Other applications for this product include instrument panel covers, camera filters, and myriad glazing uses. In these applications, the solvent and chemical resistance of the material are important.

Other allyl monomers of commercial significance are diallyl fumarate and diallyl maleate. These highly reactive trifunctional monomers contain two types of polymerizable double bonds.

Allyl methacrylate also exhibits dual functionality and finds use as both a crosslinking agent and as a monomer intermediate. Triallyl cyanurate has found use as a crosslinking agent in unsaturated polyester resins.

Most diallyl phthalate compounds are used in critical electrical/electronic applications requiring high reliability under long-term adverse environmental conditions. Compatability with modern electronic finishing technology, such as vapor phase soldering, is inherent in these materials.

E. J. Sare, PPG Industries Inc.

ALMAGEST. The name assigned by the Arabs to the great treatise on science compiled by Ptolemy during the second century. The very name Almagest, which is a hybrid combination of the Greek superlative (μϵγιοτη) with the Arabic article (al), indicates the importance of this work to the early astronomers.

The Almagest is a collection of treatises on a variety of scientific subjects. In it is to be found the complete exposition of the Ptolemaic system for the structure of the universe. Perhaps the best-known section of the Almagest is that dealing with the stars and the constellations. This section was taken from the works of Hipparchus and incorporated in the Almagest by Ptolemy, with some improvements and additions. It is in this catalogue that we first find the brightnesses of the stars divided into six magnitudes, a system that has persisted down to modern times. The positions of the stars given in the Almagest have proven of some little value in determining the constants of precession and, also, the proper motions of the stars.

ALMANAC (Astronomical). For the work of every person engaged in astronomy, whether as an astronomer in an observatory, as a navigator on a ship at sea or in the air, or as a surveyor in the field, tables of certain astronomical data are indispensable. Many, in fact most, of these tables change from year to year. Among such materials may be listed the positions of the sun, moon, and planets for every day in the year; accurate positions of stars to be used for determination of local time; tables for computing precession, nutation, aberration, etc. Such material is computed and published in almanacs several years in advance so that ships embarking on long voyages can have the data at hand when they leave port.

In addition to the ephemerides and data listed above, astronomical almanacs also contain descriptions of such phenomena as eclipses of the sun and moon, occultations of stars by the moon, eclipses and configurations of the satellites of Jupiter, etc. An examination of the preface for the American Ephemeris and Nautical Almanac or the publications in Great Britain of H.M. Stationery Office for any year will show how the work for that particular year was distributed.

ALOE (*Liliaceae*). A large genus of plants characteristic of drier parts of Africa, especially the southern part. Because of their ornamental appearance, with stiff habit and spiny-margined leaves, many of them are grown in cultivation. The rather small yellow or red flowers are borne in large masses. Many species yield from the crushed leaves a purgative juice, which is called aloes, and which has been used extensively by eastern people.

ALOPECIA (Hair Loss). Normally, this condition is confined to various portions of hair, but *alopecia universalis* designates the loss of all body hair. Alopecia is of two forms—scarring (*cicatricial*) or nonscarring (*noncicatricial*). In the scarring form, the follicles are destroyed, resulting in permanent hair loss.

Male Pattern Baldness (Androgenetic Alopecia). This condition mainly affects the vertex and frontal regions of the scalp in males. However, this form of alopecia may be seen in young women, infrequently, as the result of some underlying endocrine disease. Causes of androgenetic alopecia that have been suggested by some researchers include an excessive production of dihydrotestosterone (DHT) in the affected areas. It has been established that DHT is a potent androgenic end-organ effector in certain tissues. From time to time, cures for male baldness have been suggested. For example, there has been speculation that a specific pharmacologic blocking agent (an antiandrogen) might be applied topically to block androgenetic alopecia. Experience with animals along these lines has been encouraging, but lasting success on humans requires considerably more proof.

Alopecia Areata. This condition, still of unknown etiology, appears to be associated with a variety of probably autoimmune diseases, including pernicious anemia, thyroiditis, and Addison's disease. Complete hair loss may develop on the scalp or other hair-bearing areas, such as beard or eyebrows. The areas usually are well circumscribed. Some tenderness may accompany the lesions. In the early stage, hairs exhibit a tapered shaft and clubbed bulb, sometimes termed "exclamation point" hairs. Complete baldness in the area then generally follows. Distribution of hair loss in a peripheral band around the scalp is called *ophiasis*. Where there is no underlying endocrine disease present, a spontaneous regrowth of hair may occur within a three-year period. Other persons exhibit a more chronic course. The complete loss of scalp hair is called *alopecia totalis*.

Treatment of alopecia areata takes the form of topical application or intralesional injection of high-potency corticosteroids, which experience indicates will promote regrowth of hair, but only so long as the therapy is continued. Good temporary responses in cases of extensive alopecia areata have been obtained from systemic corticosteroids.

Drug-Induced Alopecia. Thinning of hair may result from iron deficiency anemia, hypothyroidism, or hyperthyroidism. Cytotoxic agents, including antimetabolites, such as administered in cancer therapy, can cause *anagen effluvium* by interfering with mitotic activity in the hair follicles. Hair shafts become thin and break off easily. Upon cessation of the use of cytotoxic drugs, the hair loss is reversible. Other drugs which (usually after administration for over three months) cause diffuse hair loss include heparin, triparanol, thiourea, indomethacin, lithium carbonate, nitrofurantoin, propanolol, probenecid, allopurinol, and, sometimes, excessive intakes of vitamin A.

Alopecia also has been associated with the use of oral contraceptives, the condition resembling that of androgenetic alopecia in women, as previously mentioned.

Permanent hair loss may result from certain fungal and bacterial infections and from a number of underlying causes, such as discoid lupus erythematosus, scleroderma, folliculitis decalvans, among others, the description of which is beyond the scope of this encyclopedia.

Various "treatments" for alopecia, particularly for males, have been offered to the public for a century or more with little, if any, evidence of effectiveness or universal acceptance by the medical profession. Currently, at least one pharmaceutical manufacturer is offering a product publicly, but it is obtainable only by prescription in the United States. The best source of information is a patient's dermatologist.

ALPHA CENTAURI. Ranking third in apparent brightness among the stars, Alpha Centauri has a true brightness value of 1.5 as compared with unity for the sun. Alpha Centauri is a yellow, spectral type G star and is one of the terminal stars in the pattern of the constellation Centaurus located south of the ecliptic. In the mid-1800s, the South African astronomer, Thomas Henderson, determined that Alpha Centauri is the nearest star to the sun, an estimated 4.3 light years distant. Actually, Alpha Centauri is a double star, with a third star, Proxima Centauri, revolving around the two stars. Alpha Centauri and another star in the constellation Centaurus, Beta Centauri, form a line which points quite closely to the south pole of the celestial sphere. See also **Constellations;** and **Star.**

ALPHA CHAMBER. A counter tube or counting chamber for the detection of alpha particles; often operated in the nonmultiplying (ionization chamber) or proportional region with pulse height selected to discriminate against pulses due to beta or gamma rays and to pass only those due to alpha particles.

ALPHA CRUCIS. Ranking thirteenth in apparent brightness among the stars, Alpha Crucis has a true brightness value of 4,000 as compared with unity for the sun. Alpha Crucis is a blue-white, spectral type B star and is located in the constellation Crux (Southern Cross) south of the ecliptic. Estimated distance from the earth is 400 light years. See also **Constellations.**

ALPHA CUTOFF. The frequency at which the alpha (current amplification) of a transistor has fallen to 0.7 (3 decibels) of its low-frequency value.

ALPHA DECAY. The process that occurs when alpha particles are emitted by radioactive nuclei. The name *alpha particle* was applied in the earlier years of radioactivity investigations, before it was fully understood what alpha particles are. It is known now, of course, that alpha particles are the same as helium nuclei. When a radioactive nucleus emits an alpha particle, its atomic number decreases by $Z = 2$ and its mass number by $A = 4$. The process is a spontaneous nuclear reaction.

The entire energy released by the transition is carried away by the product nuclei. Therefore, a spectrum of alpha-particle numbers as a function of energy shows a series of distinct peaks, each corresponding to a single alpha-particle transition. To conserve both energy and momentum, the energy must be shared by the two product nuclei, with the daughter nucleus ($^{A-4}Z - 2$) recoiling away from the direction of emission of the alpha particle. If E_x and M_x are, respectively, the kinetic energy and mass of the alpha particle and E_R and M_R the kinetic energy and mass of the recoiling product nucleus, the transition energy is $Q = E_\alpha + E_R$; and the kinetic energy of the emitted alpha particle is $E_\alpha = [M_R\,(M_\alpha + M_R)]Q$.

Almost all radioactive nuclides that emit alpha particles are in the upper end of the periodic table, with atomic numbers greater than 82 (lead), but a few alpha-particle emitting nuclides are scattered through lower atomic numbers. The reason why alpha-particle emitters are limited to nuclides with larger mass numbers is that generally only in this region is alpha-particle emission energetically possible. Most radioactive nuclides with smaller mass numbers emit beta-particle radiation.

See **Particles (Subatomic);** and **Radioactivity.**

ALPHA EMITTER. A radionuclide that undergoes a transformation by alpha-particle emission.

ALPHERATZ (α Andromedae). A star formerly allotted to the constellation Pegasus by the Arabs. It is situated at the northeast corner of the great square of Pegasus. The star is a spectroscopic binary with a period of approximately 100 days. See also **Constellations.**

ALPINE (Ecology). Characteristic of mountainous regions which occur between the timberline and the snowline. The term is used with reference to the flora, climate, relief, ecology, etc. of such regions. In a less restrictive sense, the term pertains to high elevations and cold climates.

ALPINE OROGENY. The term for the relatively young orogenic events which occurred in southern Europe and Asia, during which the rocks of the Alps and the remainder of the Alpine orogenic belt were strongly deformed. Most geologists restrict the era to the Tertiary, ending during the Miocene or Pliocene.

ALTAIR (α Aquilae). A star that forms, with β and γ of the constellation Aquila, a well-known line of stars sometimes referred to as the shaft of Aquila. These stars are a conspicuous feature of the early autumn sky. Ranking twelfth in apparent brightness among the stars, Altair has a true brightness value of 11 as compared with unity for the sun. Estimated distance from the earth is $16\frac{1}{2}$ light years. Altair is classified as a white star of spectral type A. See also **Constellations;** and **Star.**

ALTAZIMUTH. An instrument so mounted that it may be rotated about a horizontal and a vertical axis (i.e., rotated in altitude and azimuth); the earliest type of mounting for astronomical telescopes. Perhaps the most familiar altazimuth instrument is the ordinary surveyor's transit or theodolite.

The great advantage of the altazimuth is the ease with which it may be set up. If the instrument has been properly constructed, the horizontal and vertical axes will be strictly perpendicular to each other, and the only necessary adjustment will be to level the horizontal axis.

The altazimuth instrument is used in the field for laying down azimuth lines and for determining latitude and longitude by measuring altitudes of celestial objects. A few large, fixed altazimuth instruments are in use in observatories for accurate determination of declinations of stars, but for this purpose, the meridian circle is most commonly used.

For ordinary astronomical observing, the altazimuth instrument is less convenient than the equatorial because of the fact that the diurnal motion of the celestial sphere is parallel to the equator rather than to the horizon, with the result that the instrument has to be moved about both axes to follow the celestial objects. See also **Theodolite.**

ALTERNATING CURRENT FUNDAMENTALS. Electric currents which vary periodically with time. These currents are produced by impressing periodic voltages on an electric system. Alternating voltages are produced by rotating generators called alternators or by electronic devices called oscillators, signal generators or waveform generators. Various alternating waveforms are shown in the accompanying figure. One complete set of positive and negative values is called a *cycle*. The time T of a complete cycle is the *period* and is given in seconds. The *frequency* of a wave is the number of cycles occurring in one second. Thus f, the frequency, $= 1/T$ cycles per second. The angular frequency of the periodic wave is defined as 2π times the frequency or ω (Greek letter omega), the angular frequency, $= 2\pi f$ radians per second. For the sine wave, the instantaneous voltage $e(t) = E_m \cos(\omega t + \alpha)$. The angle α is called the phase angle and is determined by the position of the origin.

The instrument for measuring waveforms is the oscilloscope which describes a plot on the face of a cathode-ray tube. The vertical deflection of the describing electron beam is proportional to the voltage on the vertical presentation terminals and the horizontal deflection is proportional to the linear sweep-voltage on the vertical presentation terminus. When the wave being observed is repetitive at a fast enough rate, the display appears as a steady figure.

The average value of the sine wave voltage for a complete cycle is 0. The average value for half a cycle is $2E_m/\pi$. The rms (root-mean-square) value of the sine wave voltage is $\sqrt{2}E_m/2 = 0.707E_m$. See Fig. 1.

Through common usage, the expression *alternating current* or ac refers to a current or a voltage that has a waveform closely approximating the sine wave. There are many advantages to the use of sinusoidal voltages such as:

1. Efficient transmission of electrical energy over long distances by stepping up the voltage with static transformers.
2. Simplification in design and reduction in cost of generators and motors.
3. Basic waveform for signal processing in the generation of modulated carriers for communication systems.
4. Signal analysis using the Fourier methods. See **Frequency (Electric) Measurement**.
5. Systems analysis in terms of frequency and phase response. See also **Oscilloscope; Signal Generator**.

Alternating Current Circuits

A circuit is an interconnection of electrical elements consisting of energy sources, resistors, capacitors, and inductors having self and mutual inductance. The steady-state response of a circuit or electrical network to sinusoidal sources obtained in complex representation. The complex form of a sinusoidal voltage

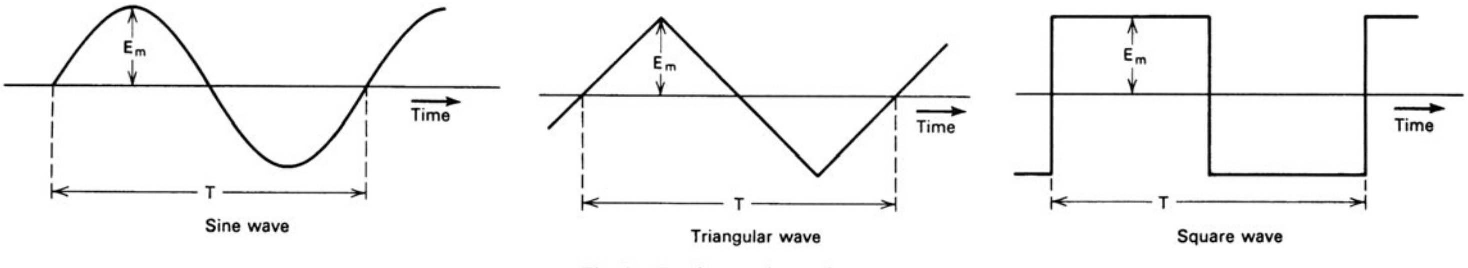
$$e(t) = E_m \cos(\omega t + \alpha)$$

is

$$\begin{aligned} E_m \epsilon^{j(\omega t + \alpha)} &= (E_m \epsilon^{j\alpha})(\epsilon^{j\omega t}) \\ &= \mathbf{E}_m \epsilon^{j\omega t}, (j = \sqrt{-1}) \end{aligned}$$

The complex $\mathbf{E}_m = E_m \epsilon^{j\alpha} = E_m \angle \underline{\alpha}$ is called the phasor voltage in terms of its maximum value. $\mathbf{E} = \mathbf{E}_m \sqrt{2} = 0.707E_m \angle \alpha$, the phasor voltage in terms of its rms or effective value, is called simply the phasor voltage. It can be seen that the instantaneous voltage is the real part of its complex representation. The complex representation of the sinusoidal current

$$i(t) = I_m \cos(\omega + \beta) \text{ is } I_m \epsilon^{j(\omega t + \beta)} = (I_m \epsilon^{j\beta}) \epsilon^{j\omega t} = \mathbf{I}_m \epsilon^{j\omega t}$$



Sine wave

Triangular wave

Square wave

Fig. 1. Fundamental waveforms.

where $\mathbf{I}_m$, the phasor current in terms of its maximum, is $I_m\epsilon^{j\beta} = I_m \angle \beta$. The phasor current $\mathbf{I} = \mathbf{I}_m/\sqrt{2} = 0.707I_m \angle \beta$. Again, the instantaneous value of the current $i(t)$ = real part of $\mathbf{I}_m\epsilon^{j\omega t}$. Phasors are expressed in either polar form, magnitude and angle, or rectangular form, real and imaginary parts.

For an inductance, the voltage

$$e(t) = L\frac{di(t)}{dt}$$

Now $i(t)$ in complex representation is $\mathbf{I}_m\epsilon^{j\omega t}$, so $e(t)$ in complex representation is $\mathbf{I}_m(j\omega L)\epsilon^{j\omega t}$ and

$$\mathbf{E}_m = \mathbf{I}_m(j\omega L)$$

The ratio of the two phasors,

$$\frac{\mathbf{E}_m}{\mathbf{I}_m} = \frac{\mathbf{E}}{\mathbf{I}} = j\omega L = \omega L \angle \frac{\pi}{2} = (\omega L)\epsilon^{j\omega/2}$$

The instantaneous voltage $e(t) = E_m \cos(\omega t + \alpha)$ = real part of

$$\mathbf{I}_m(\omega t)\epsilon^{j(\omega t + \pi/2)} = \mathbf{I}_m(\omega L)\cos\left(\omega t + \beta + \frac{\pi}{2}\right)$$

So

$$E_m = I_m\,(\omega L)$$

and

$$\alpha = \beta + \frac{\pi}{2}$$

Figure 2 shows the circuit diagram, the sinusoidal waveforms and the phasor diagram. As $E_m = I_m(\omega L)$ then $E_{rms} = I_{rms}(\omega L)$ or the rms value of the voltage across an inductance is equal to $(\omega L) = X_L$, the inductive reactance (given in ohms) multiplied by the rms value of the current through the inductance. The sinusoidal current lags the voltage by $\pi/2$ radians.

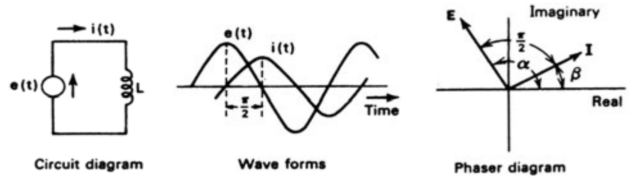


Fig. 2. Case where sinusoidal current lags the voltage by $\pi/2$ radians.

For a capacitance, the current

$$i(t) = C\frac{de(t)}{dt}$$

In complex representation,

$$\mathbf{I}_m\epsilon^{j\omega t} = \mathbf{E}_m(j\omega C)\epsilon^{j\omega t}$$

or

$$\mathbf{I}_m = \mathbf{E}_m(j\omega C)$$

$$\frac{\mathbf{E}_m}{\mathbf{I}_m} = \frac{\mathbf{E}}{\mathbf{I}} = \frac{1}{j\omega C} = -\frac{j}{\omega C} = jX_C = \frac{1}{\omega C}\angle\frac{\pi}{2}$$

The instantaneous voltage $e(t) = I_m(1/\omega C)\cos(\omega t + \beta - \pi/2)$. Therefore $\alpha = \beta - \pi/2$ and $E_{rms} = I_{rms}(-X_C)$, where X_C, the capacitive reactance, is $-1/\omega C$, or the rms value of the voltage across a capacitor is equal to the magnitude of the capacitive reactance multiplied by the rms current. The sinusoidal current leads the voltage by $\pi/2$ radians. Figure 3 illustrates this case.

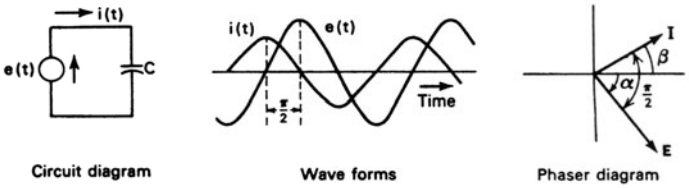


Fig. 3. Case where sinusoidal current leads the voltage by $\pi/2$ radians.

For a resistance, the voltage, $e(t) = Ri(t)$. In the complex representation,

$$\mathbf{E}_m\epsilon^{j\omega t} = R\mathbf{I}_m\epsilon^{j\omega t}$$

or

$$\mathbf{E}_m = \mathbf{I}_m R, \quad \frac{\mathbf{E}_m}{\mathbf{I}_m} = \frac{\mathbf{E}}{\mathbf{I}} = R$$

The instantaneous voltage $e(t) = I_m R\cos(\omega t + \beta)$, where $E_m \div I_m R$ and $\alpha = \beta$.

The rms value of the voltage across a resistor is equal to the product of the current through the resistance and the value of the resistance. The sinusoidal current is in phase with the voltage. Figure 4 illustrates this case.

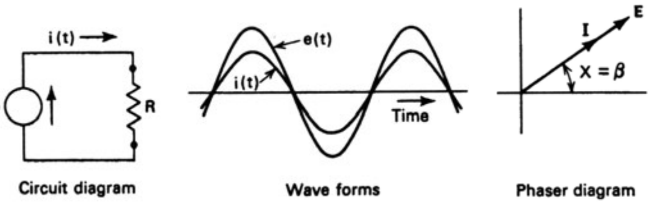


Fig. 4. Case where sinusoidal current is in phase with the voltage.

The responses of a network to sinusoidal sources are determined directly from the general but simple problem of the steady-state response of a network to sinusoidal sources.

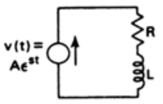


Fig. 5. Circuit with an exponential voltage source.

Thus for the circuit shown in Fig. 5 with an exponential voltage source $A\epsilon^{st}$, the application of Kirchhoff's laws gives

$$Ri + L\frac{di}{dt} = A\epsilon^{st}$$

As the particular integral is the steady state solution,

$$i(t) = \frac{A\epsilon^{st}}{R + sL}$$

For sinusoidal response, using complex representation

$$A \text{ is } \mathbf{E}_m, \quad s = j\omega$$

and

$$i(t) \text{ is } \mathbf{I}_m\epsilon^{j\omega t}$$

The equation for the current to the network becomes

$$\mathbf{I}_m\epsilon^{j\omega t} = \frac{\mathbf{E}_m\epsilon^{j\omega t}}{R + j\omega L}$$

or

$$\mathbf{I}_m = \frac{\mathbf{E}_m}{R+j\omega L}, \quad \mathbf{I} = \frac{\mathbf{E}}{R+j\omega L} = \frac{\mathbf{E}}{\mathbf{Z}} = \mathbf{EY}$$

The impedance of the network (sometimes referred to as the complex impedance) is

$$\mathbf{Z} = \frac{\mathbf{E}_m}{\mathbf{I}_m} = \frac{\mathbf{E}}{\mathbf{I}}$$

For this network, the impedance is $R + j\omega L$ or its real part is the Resistance R and its imaginary part, the inductive reactance ωL. The admittance of the network, $\mathbf{Y}$, is equal to $\mathbf{I}_m/\mathbf{E}_m = \mathbf{I}/\mathbf{E} = 1/\mathbf{Z}$. For this network, the admittance is $1/(R = j\omega L)$. As

$$\mathbf{I}_m = \frac{\mathbf{E}_m}{R+j\omega L} = \frac{E_m \angle \alpha}{\sqrt{R^2+(\omega L)^2}\angle \theta} = \frac{E_m \angle(\alpha-\theta)}{\sqrt{R^2+(\omega L)^2}}$$

So the instantaneous current

$$i(t) = \frac{E_m}{\sqrt{R^2+(\omega L)^2}}\cos(\omega t+\alpha+\theta), \quad \theta = \tan^{-1}\frac{\omega L}{R}$$

$$= I_m \cos(\omega t + \beta)$$

$$I_m = \frac{E_m}{\sqrt{R^2+(\omega L)^2}}$$

or

$$I_{rms} = \frac{E_{rms}}{\sqrt{R^2+(\omega L)^2}}$$

and $\beta = \alpha - \theta$ or the current lags the voltage by the angle θ. Figure 6 shows the waveform and the phasor diagram.

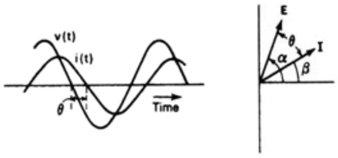


Fig. 6. Waveform and phasor diagram for case of Fig. 5.

For the series R-L-C circuit,

$$Ri + L\frac{di}{dt} + \frac{1}{C}\int i\,dt = A\epsilon^{st}$$

The particular integral is

$$i(t) = \frac{A\epsilon^{st}}{R + sL + \frac{1}{sC}}$$

This relation may be written directly from the network, labelling each element as shown in Fig. 7 and calculating the current as if it were a dc circuit with each element equivalent to a resistor. For the sinusoidal response

$$\mathbf{I} = \frac{\mathbf{E}}{R + j\omega L + 1/j\omega C} = \frac{\mathbf{E}}{R + j(X_L + X_C)} = \frac{\mathbf{E}}{\mathbf{Z}}$$

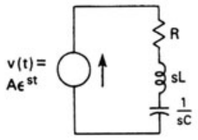


Fig. 7. Example of labeling each element in network.

and

$$\mathbf{Z} = R + j(X_L + X_C) = \sqrt{R^2 + (X_L + X_C)^2}\angle\theta$$

where

$$\theta = \tan^{-1}\frac{X_L + X_C}{R}$$

and

$$i(t) = \frac{E_m}{\sqrt{R^2 + (X_L + X_C)^2}}\cos(\omega t + \alpha - \theta)$$

So

$$I_{rms} = \frac{E_{rms}}{\sqrt{R^2+(X_L+X_C)^2}} \quad \text{and} \quad \beta = \alpha - \theta$$

When $|X_L| > |X_C|$, θ is positive and the current in the circuit lags the impressed voltage, and when $|X_L| < |X_C|$, θ is negative and the current leads the impressed voltage.

For the parallel R-L-C circuit with a current source as shown in Fig. 7, the dual of circuit shown in Fig. 8,

$$\frac{e}{R} + C\frac{de}{dt} + \frac{1}{L}\int e\,dt = A\epsilon^{st}$$

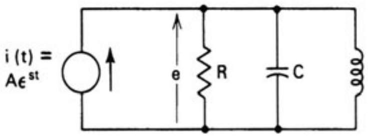


Fig. 8. Parallel R-L-C circuit with a current source.

The particular integral is

$$e(t) = \frac{A\epsilon^{st}}{\frac{1}{R} + sC = \frac{1}{sL}}$$

In complex representation,

$$A = \mathbf{I}_m, \qquad s = j\omega$$

and $e(t)$ is $\mathbf{E}_m\epsilon^{j\omega t}$. So

$$\mathbf{E}_m = \frac{\mathbf{I}_m}{\frac{1}{R} + j\omega C + \frac{1}{j\omega L}} = \frac{\mathbf{I}_m}{\mathbf{Y}}$$

or

$$\mathbf{E} = \frac{\mathbf{I}}{\frac{1}{R} + j\left(\omega C - \frac{1}{\omega L}\right)} = \frac{\mathbf{I}}{G + j(B_C + B_L)} = \frac{\mathbf{I}}{\mathbf{Y}}$$

The admittance $\mathbf{Y} = G + j(B_C + B_L)$ where

G = conductance = $1/R$

B_C = capacitive susceptance = ωC

B_L = inductive susceptance = $-\frac{I}{\omega L}$

$$e(t) = \frac{I_m}{\sqrt{G^2 + (B_C + B_L)^2}}$$

$$\theta = \tan^{-1}\frac{B_C + B_L}{G}$$

The techniques for solving ac network problems are exactly the same as those for networks with exponential sources but with $j\omega$ replacing s. For impedances in series

$$Z_t = \sum_{i=1}^{n} Z_i$$

and for admittances in parallel,

$$Y_t = \sum_{i=1}^{n} Y_i$$

Also, the same mesh and nodal analysis may be applied as in dc circuits with resistances and conductances being replaced by impedances and admittances in complex representation.

The response of a linear network to a periodic source function that is not sinusoidal may be determined by describing the function as a sum of sine waves in a Fourier analysis. The sum of responses to each sine wave is the total response.

The instantaneous power supplied to an ac circuit is $p = ei$, and the average power is

$$P_{av} = \frac{1}{T}\int_0^T ei\,dt$$

ALTERNATING GRADIENT FOCUSING. This principle, discovered independently by Christofilos in 1950 and by Courant, Livingston, and Snyder in 1952, is based on the observation from geometrical optics that a combination of a convex lens and a concave lens of equal focal length have a net focusing effect. This type of focusing of moving charged particles is accomplished by appropriately shaped magnetic fields. But, since magnetic fields cannot focus simultaneously in both the vertical and horizontal planes that can be constructed through a line representing the direction of motion of a beam of charged particles, alternate lenses that focus first in the horizontal direction and then in the vertical direction are needed. This can be accomplished for an external beam of charged particles by pairs of quadrupole magnets. The same principle has been applied to the magnetic fields in particle accelerators. In a synchrotron, for example, a section of magnet focuses axially and defocuses radially, while the following section defocuses axially and focuses radially.

ALTERNATING GROUP. The group of even permutations of *n* objects. See **Permutation.**

ALTERNATING SERIES. A series whose terms are alternately positive and negative. See **Series.**

ALTERNATION OF GENERATIONS. A term used to describe the alternation of two distinct body forms, one reproducing sexually and one reproducing asexually, in the life cycle of a plant or animal. A typical example is to be found in the hydrozoan coelenterate, *Obelia*. See Fig. 1. One generation, the hydroid stage, reproduces asexually by budding and forms a colony consisting of a stalk bearing many polyps. As the colony grows older, reproductive polyps are formed which produce buds that grow into small jellyfish known as medusae. Some polyps produce male and others produce female medusae. These medusae break free of the polyp and swim in the water. The males produce sperms which swim to the eggs produced by the females. The zygote grows into a ciliated larva which swims around for a time and then settles down and becomes attached. It then grows into a hydroid stage.

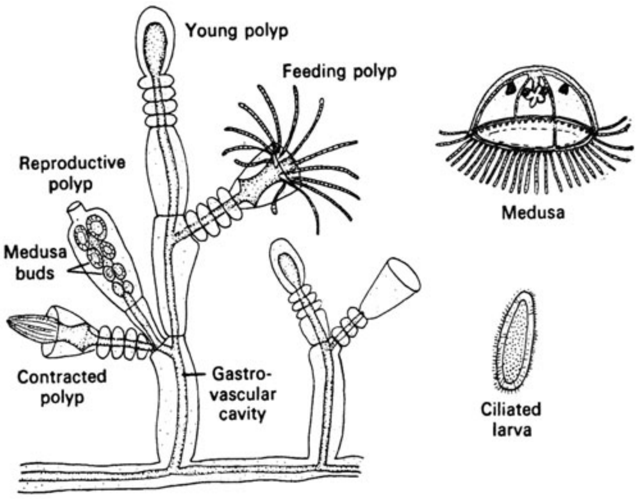


Fig. 1. The life cycle of *Obelia*.

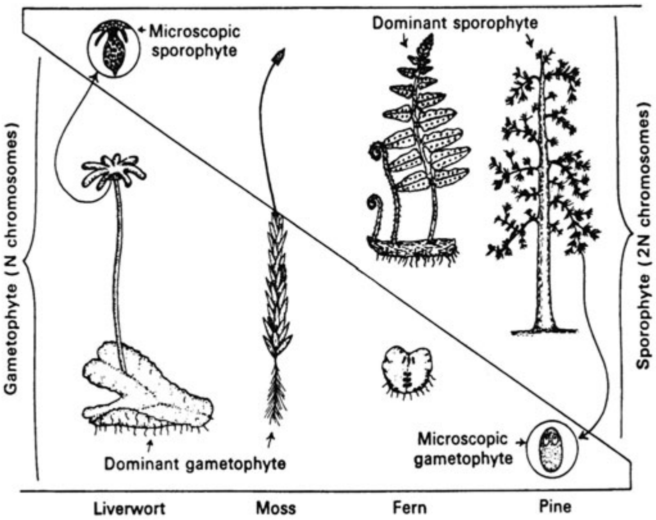


Fig. 2. Alternation of generations is demonstrated by the gradual shift of emphasis from the dominant gametophyte generation (exemplified by the liverworts) to the dominant sporophyte in the seed plants (exemplifed by the pine).

Parasitic flatworms, such as flukes and tapeworms, and the malarial parasite (*Plasmodium*) are among the other animals that have alternation of generations.

In all except some of the simplest plants there is alternation of generations. The sporophyte generation reproduces asexually by means of spores. These spores are produced by a series of two special cell divisions known as meiosis which reduces the chromosome number to half, the haploid number. A spore grows into a gametophyte generation which will have the haploid number. Gametes are produced by the gametophyte, and when these unite a zygote is formed with the diploid number. The zygote then grows into a sporophyte generation.

As plants become larger and more complex, there is a tendency toward a reduction in the size of the gametophyte generation. In liverworts, the sporophyte is microscopic in size; in the mosses the gametophyte is still dominant, but there is a clearly visible sporophyte; in the ferns the sporophyte is dominant, but the gametophyte is still present as a distinct and separate plant; in the seed plants the gametophyte is microscopic in size. See Fig. 2.

ALTERNATOR. An electromotive force is generated in a conductor when it is moved so as to cut the lines of force between the poles of a magnet. The elementary principle of a simple two-pole, single-phase alternator is shown in Fig. 1. When the magnet revolves it will carry with it lines of force which will cut the conductor, which is a wire loop embedded in the stationary portion called the armature, and will generate ac.

This elementary principle must be expanded in several directions if a practical generator of ac is to be had. First, the rotating part, or rotor, must have magnetic strength in excess of that which could be obtained from a simple permanent magnet. In other words, the poles must be formed by electromagnets whose excitation, in the form of dc, must be carried to the rotor through slip-ring connections. The rotor is called the

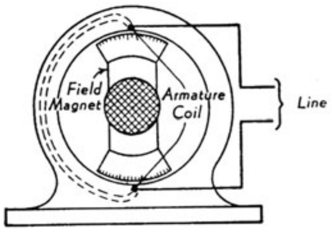


Fig. 1. Elementary alternator.

field, and the current it uses is called the field current. In high-speed steam turbine-driven alternators as few as two poles are often used, while in slow-speed water-turbine units the number is frequently nearly 100. The stationary part, called the stator, or armature, usually has three sets of overlapping coils, connected in three separate circuits. These three circuits, or phases, are usually connected in one or the other ways shown in Fig. 2. The *Y* connection is preferred because of the usefulness of the neutral point, and the fact that the line voltage is $\sqrt{3}$ times the phase voltage, whereas it is only equal to the phase voltage in Δ connection. The neutral point is connected to the fourth wire of a four-wire, three-phase system, and left unconnected, or grounded, in the three-wire system. Several advantages are realized by making the rotating part the field, and the stationary part the armature. The ac may be generated at very high voltages because it is not necessary to connect it through movable contacts as would be the case if the armature revolved. It is not necessary to conduct high-load currents through slip rings and brushes if the armature is fixed. The armature conductors can be very rigidly braced in position, and may be much better disposed than if they were required to be in the rotor.

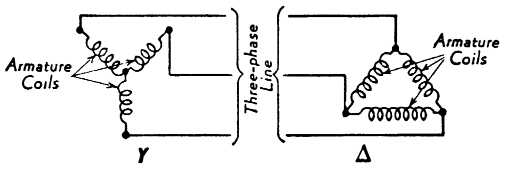


Fig. 2. Comparison of *Y* and *Delta* connections of three-phase alternator windings.

Engine and hydraulic turbine-driven alternators are in the slow-speed class, and are characterized by large diameter, short length, and many poles. The steam turbine-driven alternator is a high-speed machine having a length larger than its diameter. Standard speeds of turbine-driven alternators range from 1200 to 3600 rpm, with 1800 rpm very common practice.

Basically, the alternator is a device for converting mechanical into electrical energy. While it is able to do this with a high degree of efficiency, it does suffer the following losses:

1. Friction and windage from bearings, brushes, and fan action of the rotor.
2. Core loss, which is the result of eddy currents and hysteresis in the iron core.
3. Resistance heating loss in the armature and field conductors.
4. Resistance loss of the field rheostat.
5. Exciter loss.
6. Ventilation loss.

Practically all generator losses appear as heat in and about the windings, and to maintain these at a safe working temperature, a cooling medium must be employed. Air has been the medium generally used. The rotor may or may not be able to produce its own fan action, depending on the size, speed, and construction of the rotor. Ventilating air frequently has to be brought through a duct and discharged through the alternator. Hydrogen is rapidly becoming a common cooling medium as it is much more effective and produces less windage loss. For its use the generator must be totally enclosed and gas-tight. To supply the dc for the field, a source of dc at 110–250 volts is necessary. This is delivered from a small dc generator called the exciter. The exciter may be driven from an extension of the alternator shaft, or it may be driven by some independent means, such as motor or engine.

When alternators are operated in parallel, the division of load between them is not accomplished by changing the generated voltage, as in dc generators, but by changing power input from the prime mover through the adjustment of the engine or turbine governor. Before being paralleled, two alternators must have the same phase sequence and frequency. Their voltages must be equal, and in phase; that is, with the peaks of wave forms coincident in time and direction.

ALTIMETRY. The technique of using air pressure to measure altitude. In altimetry there is a precisely ordered atmosphere called the *Standard Atmosphere* in which there is an exclusive value of altitude associated with a given value of air pressure. An altimeter converts static pressure into altitude. It is this precise and mathematically rigid relationship between static air pressure and altitude that the altimeter incorporates into its mechanism.

The standard atmosphere is defined by the following parameters:

(1) Zero pressure altitude corresponds to:

29.9213 inches of mercury
1013.250 millibars
760.000 millimeters of mercury

(2) Temperature at zero pressure altitude is 15°C which is equivalent to 288.16 K.
(3) Density of air at zero pressure altitude is 0.0012250 gram per cubic centimeter.
(4) The rate of temperature decrease with altitude is 6.5° per kilometer from zero pressure altitude up to 11 kilometers.
(5) The temperature at 11 kilometers is −56.5°C.
(6) The temperature from 11 kilometers up to 25 kilometers is constant at −56.5°C.
(7) The acceleration of gravity is everywhere 980.665 centimeters/sec^2.*
(8) There is no water vapor in the physical make-up of the atmosphere.*
(9) There is perfect obedience to the perfect gas law and the hydrostatic equation.*

In this concept of standard atmosphere, there is a unique value of density, pressure, and temperature at each altitude. The relationship between pressure and height is given by:

$$H = \frac{T_0}{\beta}\left[1 - \left(\frac{P_0}{P}\right)^{R\beta/g}\right]$$

where P_0 = the reference pressure which *in practice* is usually one of the following:

(1) Mean sea level pressure
(2) Field pressure
(3) 29.92 inches of mercury (1013.3 millibars)
T_0 = the standard atmosphere temperature at P_0
β = the temperature rate of decrease with altitude
R = the gas constant for standard air
g = acceleration due to gravity
P = the sensed and measured pressure by the altimeter
H = the altitude corresponding to P in the standard atmosphere

Millibars of pressure are used in much of international aviation instead of inches of mercury, but the latter is still the common usage in much of North America.

Toussaint's Formula. A rule proposed by Toussaint for the linear decrease of temperature with height in the atmosphere for which the temperature at mean sea level is 15°C is given by:

$$t = 15 - 0.0065z$$

where t is the temperature in degrees centigrade and z is the geometric height in meters above mean sea level. Toussaint's formula is used to determine temperature below 11,000 meters in the ICAO standard atmosphere.

Field Pressure. This is the atmospheric pressure measured and observed at the elevation of the airport or weather observatory in the real atmosphere by use of a mercurial barometer or a calibrated aneroid instrument. All pressure-related functions in altimetry must start with field pressure.

---

*It should be stressed that the foregoing statements are part of the arbitrary, but important definition of standard atmosphere and should be interpreted in this context. For a more detailed definition of standard atmosphere, see **Atmosphere (Earth).**

Field pressure *reduction to mean sea level* relates a calculated atmospheric pressure at sea level to field pressure by application of standard atmosphere relations. Field elevation above mean sea level is known and field pressure is measured. The pressure increment dictated by the standard atmosphere is added to the field pressure to obtain a value for mean sea level pressure.

Mean Sea Level Pressure. This is used extensively in altimeter-measured altitudes above sea level up to 18,000 feet (5486 meters). Mean sea level pressure is also known as *altimeter setting*, to which altimeters in airborne aircraft are set while flying below 18,000 feet (5486 meters).

Pressure Altitude. This is height in the standard atmosphere measured positively upward from the zero elevation of 29.92 inches of mercury (1013.3 millibars). An altimeter carrying a reference pressure of 29.92 inches of mercury (1013.3 millibars) will display pressure altitude.

Altimeter. A very sensitive instrument that converts static air pressure to altitude according to the precise relationship of the pressure and height in the standard atmosphere. There is a *reference pressure* adjustable setting on the altimeter that can be set to any reference pressure within the operating range of the altimeter. The altimeter measures altitude positively upward (negatively downward) from the level of the reference setting. In practice, three reference settings are predominantly used:

(1) *Reference setting at field pressure* will cause the altimeter to measure altitude positively upward from the level of the airport. When sitting on the airport, an aircraft altimeter will read zero.

(2) *Reference setting at mean sea level pressure*, also called "altimeter setting," will cause the altimeter to read altitude above mean sea level.

(3) *Reference setting at 29.92 inches of mercury* (1013.3 millibars) will cause the altimeter to measure altitude positively upward from the level where a pressure of 29.92 inches of mercury (1013.3 millibars) prevails in the real atmosphere. Altitude readings obtained with the reference set at 29.92 inches of mercury (1013.3 millibars) are known as "pressure altitude."

The altimeter may be an aneroid barometer calibrated to read in feet of altitude instead of barometric pressure. It thus reads pressure, not density, altitude. The heart of this instrument is a circular evacuated capsule, one face of which is fastened to the case, while the other is attached through a system of multiplying levers (terminating in a chain wrapping around the spindle) to a spindle carrying a pointer which moves over the scale. A U-shaped spring holds the capsule distended and atmospheric pressure tends to collapse it.

When mounted in an airplane which is climbing to higher altitudes, the capsule is subjected to decreasing atmospheric pressure, allowing the spring to distend the capsule and move the pointer. A small hairspring keeps the mechanism taut and the pointer will move upon the slightest movement of the capsule face in either direction.

The instrument may be made to indicate altitude above any given point, e.g., the airport runway, by rotating the dial containing the scale until the pointer registers with zero altitude, this being done when the airplane is on the runway. A knob having a pinion engaging a gear on the periphery of the dial makes this adjustment.

Vertical Separation. The vertical separation of airborne aircraft is accomplished by use of onboard altimeters and the application of altimetry.

(1) Below 18,000 feet (5486 meters), all aircraft have their altimeters set to a common altimeter setting (mean sea level) in a given traffic control area. If there are deviations from the real altitude above sea level, all aircraft carry the same deviation.

(2) Above 18,000 feet (5486 meters), all aircraft use a reference setting of 29.92 inches of mercury (1013.3 millibars) and operate on pressure altitude.

Terrain clearance when using altimeter-indicated altitude requires some precaution in air that is colder than the standard atmosphere. Mountain peaks have fixed real heights which obviously must be exceeded.

(1) If the reference setting is mean sea level pressure, the altimeter indicates altitude above sea level. When the real atmosphere is colder than the standard atmosphere, the altimeter will display a reading *too high*, i.e., the aircraft will have *less* real altitude than shown on the altimeter.

(2) If the reference setting is 29.92 inches of mercury (1013.3 millibars), the altimeter indicates altitude above the level where this value of *P* prevails. In addition to the cold air deviation, an additional caution is needed. When sea level pressure is less than 29.92 inches of mercury (1013.3 millibars), the altimeter starts counting altitude from *below sea level*. The altitude displayed by the altimeter will be too high by reason of the below-standard sea level pressure; and additionally may be in error because the air is colder than standard air.

(*Mean Sea Level*)—The average height of the sea surface, based upon hourly observation of the tide height on the open coast or in adjacent waters that have free access to the sea. These observations are to have been made over a "considerable" period of time. In the United States, mean sea level (MSL) is defined as the average height of the surface of the sea for all stages of the tide over a 19-year period.

(*Mean Sea Level Pressure*)—The value assigned to surface pressure of the standard atmosphere; 29.92 inches of mercury; or 1013.3 millibars.

Other Forms of Altimeters. A radio altimeter is an electronic instrument that may operate on one of the following principles: (1) pulse radar techniques which measure height in terms of the transit time of the radar pulse; (2) continuous-wave radar which measures height in terms of the phase difference between the transmitted and received signals; and (3) the variation of the electrical capacitance between aircraft and ground.

For references see entries on **Climate;** and **Meteorology.**

Peter E. Kraght, Certified Consulting Meteorologist,
Mabank, Texas.

ALTITUDE. The term *altitude* is used synonymously for height, or distance above the surface of the earth. The altimeter is used in aviation to measure the altitude of a plane above sea level.

In astronomy and navigation the altitude of a celestial object is that coordinate in the horizontal system of spherical coordinates that is measured in the plane of the vertical circle through the object from the horizon to the object. Altitude is probably the most frequently measured of all celestial coordinates since it is universally used in celestial navigation to find lines of position for the determination of the location of a ship at sea, or an aircraft in the air. Altitude is also measured by geodetic surveyors for accurate determination of latitude and is used by astronomers for determination of the declinations of the stars.

The principal points of high and low elevation on the earth's surface are given in the entry on **Earth.**

ALUM. A series (*alums*) of usually isomorphous crystalline (most commonly octahedral) compounds in which sulfate is usually the negative ion. Alums have the general formula $M^{I}M^{III}(AX_4)\cdot 12H_2O$. The first is potassium or a higher alkali metal (sodium alums are rare) or thallium(I) or ammonium (or a substituted ammonium ion) and the second is a tripositive ion of relatively small ionic radius (0.5–0.7Å). The tripositive ions of larger ionic radius such as the Lanthanides form double sulfates, but not alums. The tripositive ions in alums, in roughly the order of number of known compounds, are aluminum, chromium, iron, manganese, vanadium, titanium, cobalt, gallium, rhodium, iridium, and indium.

The term *alum*, itself, refers to potassium alum, potassium aluminum sulfate, $KAl(SO_4)_2\cdot 12H_2O$. Other common alums are ferric ammonium alum, $NH_4Fe(SO_2)_2\cdot 12H_2O$, and sodium chrome alum, $NaCr(SO_4)_2\cdot 12H_2O$.

ALUMINA. See **Adsorption Operations; Bauxite.**

ALUMINUM. Chemical element symbol Al, at. no. 13, at. wt. 26.98, periodic table group 13, mp 660 ± 1°C, bp 2452 ± 15°C, sp gr 2.699. In Canada and several other English-speaking nations, the spelling of the element is *aluminum*. The element is a silver-white metal, with bluish tinge, capable of taking a high polish. Commercial aluminum has a purity of 99% (minimum), is ductile and malleable, possesses good weldability, and has excellent corrosion resistance to many chemicals and most common substances, particularly foods. The electrical conductivity of aluminum (on a volume basis) is exceeded only by silver, copper, and gold. First ionization potential, 5.984 eV; second, 18.823 eV; third, 28.44 eV. Oxidation potentials, $Al \rightarrow Al^{3+} + 3e^-$, 1.67 V; $Al + 4OH^- \rightarrow AlO_2^- + 2H_2O + 3e^-$, 2.35 V. Other important physical properties of aluminum are given under **Chemical Elements.**

Aluminum occurs abundantly in all ordinary rocks, except limestone and sandstone; is third in abundance of the elements in the earth's crust (8.1% of the solid crust), exceeded only by oxygen and silicon, with which two elements aluminum is generally found combined in nature; present in igneous rocks and clays as aluminosilicates; in the mineral cryolite in Greenland as sodium aluminum fluoride Na_3AlF_6; in the minerals corundum and emery, the gems ruby and sapphire, as aluminum oxide Al_2O_3; in the mineral bauxite in southern France, Hungary, Yugoslavia, Greece, the Guianas of South America, Arkansas, Georgia, and Alabama of the United States, Italy, Russia, New Zealand, Australia, Brazil, Venezuela, and Indonesia as hydrated oxide $Al_2O(OH)_4$; in the mineral alunite or alum stone in Utah as aluminum potassium sulfate $Al_2(SO_4)_3 \cdot K_2SO_4 \cdot 4Al(OH)_3$. See also **Bauxite;** and **Cryolite.**

Uses. Because of its corrosion resistance and relatively low cost, aluminum is used widely for food-processing equipment, food containers, food-packaging foils, and numerous vessels for the processing of chemicals. Because of the presence of aluminum in soils and rocks, there are natural traces of the element in nearly all foods. The processing of foods in copper vessels will cause destruction of vitamins, whereas aluminum does not accelerate the degradation of vitamins. Aluminum foil has been used to cover severe burns as a means to enhance healing. Because of its good electrical conductivity (exceeded only by gold, silver and copper), aluminum is used as an electrical conductor, particularly for high-voltage transmission lines. For the same conductance, the weight of aluminum required is about one-half that of annealed copper. The greater diameter of aluminum conductors also reduces corona loss. But, because aluminum has about 1.4 × the linear temperature coefficient of expansion as compared with annealed copper, the changes in sag of the cable are greater with temperature changes. For long spans requiring high strength, the center strand may be a steel cable, or supporting steel cables may be used. Aluminum is used for bus bars because of its large heat-dissipating surface available for a given conductance.

Aluminum alloys readily with copper, manganese, magnesium, silicon, and zinc. Many aluminum alloys are commercially available. See also **Aluminum Alloys.**

Because aluminum, particularly in alloys, combines strength with light weight, it has been a favorite construction material for transportation equipment, such as airplanes, the early dirigibles, and parts of rail cars. With the emphasis on fuel economy during the past decade or two, the automotive industry has become a major user of aluminum alloys. For example, important developments include cast-aluminum engine blocks with metallurgically bonded-in-place cylinder liners, cast-spun wheels, aluminum space frames, and auto radiators. The first mass production of aluminum radiators occurred in the 1980s (Ford Motor Co.) and it is predicted that nearly all Ford vehicles will feature aluminum radiatiors by the mid-1990s. Of course, other auto firms worldwide are also actively pursuing the metal because of its light weight and other attractive properties.

The applications of aluminum are growing larger and more varied with the introduction of advanced materials engineering techniques, including sandwich-type construction, laminates, aluminum powder metallurgy, and composite technology.

Discovery and Early Production. Although aluminum was predicted by Lavoisier (France) as early as 1782, when he was investigating the properties of aluminum oxide (alumina), the metal was not isolated until 1825 by H. C. Oersted (Denmark). Oersted obtained an impure aluminum metal by heating potassium amalgam with anhydrous aluminum chloride, followed by distilling off the mercury. Using similar methods, Woehler (Germany) produced an aluminum powder in 1827. In 1854, Deville (France) and Bunsen (Germany) separately but concurrently found that aluminum could be isolated by using sodium instead of potassium in the amalgam of Oersted and Woehler. Deville exhibited his product at the Paris Exposition of 1855, after which Napoleon III commissioned Deville to improve the process and lower the cost. Because of improved processing techniques, the price declined from \$115 to \$17 per pound (\$254 to \$38 per kilogram) by 1859, with numerous plants throughout France. The price was further reduced to \$8 per pound (\$17.60 per kilogram) by 1885. The first breakthrough for production of aluminum on a large scale and at a much lower cost occurred as the result of experimentation by Hall (Oberlin, Ohio) who found that metallic aluminum could be produced by dissolving Al_2O_3 (alumina) in molten $3NaF \cdot AlF_3$ (cryolite) at a temperature of above 960°C and then passing an electric current through the bath. Heroult (France) independently discovered the same process.

Contemporary Production Methods. For many years the electrolytic process has dominated. The major difference found in aluminum production plants relate to the design of the electrolytic cells. Basically, the cells are large carbon-lined steel boxes. The carbon lining serves as the cathode. Separate anodes are immersed in the bath. The electric circuit is completed by passage of current through the bath. Thus, the alumina is decomposed into aluminum and oxygen by action of the electric current. The molten aluminum collects at the bottom of the cell from which it is siphoned off periodically. The released oxygen combines with carbon at the electrodes to form CO_2. The bath is replenished with alumina intermittently. The carbon anodes must be replaced periodically. Additional cryolite also is required to make up for volatilization losses. The several reactions which occur during electrolysis still are not fully understood.

Energy requirements for aluminum production in modern plants approximate 6 to 8 kilowatt-hours per pound (13.2 to 17.6 kilowatt-hours per kilogram). Per weight unit of aluminum produced, 0.4 to 0.6 weight units of carbon, 1.9 weight units of alumina, and 0.1 weight unit of cryolite are required. One form of electrolytic cell is shown in Fig. 1.

An estimate (1987) showed that the aluminum industry accounts for nearly 1.5% of the annual world energy consumption. Because of the need to conserve energy and concurrently to alleviate (even to a small extent) the disposal of solid waste materials, a program was initiated in some industrial nations to collect and recycle the aluminum cans used mainly by the carbonated beverage industry. The program has proved to be more successful than originally predicted. As of 1990, over one-half

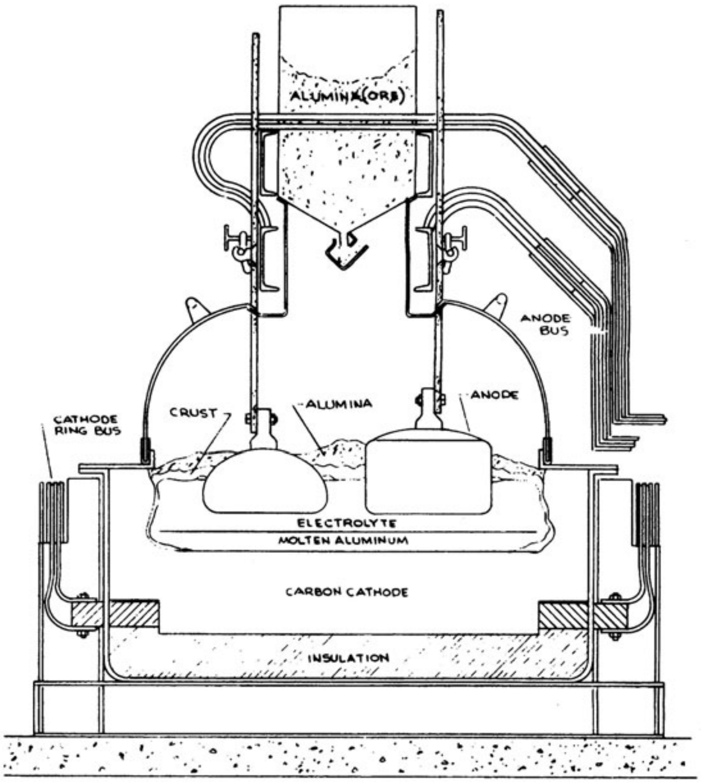


Fig. 1. Aluminum electrolytic reduction cell of the prebake type. The anodes are constructed of separate blocks of carbon which have been prebaked. There is a lead to the main bus bar from each block. (*Reynolds Metals Company.*)

Fig. 2. Recycling of aluminum beverage cans and other scrap saves nearly 95% of the energy required when producing the virgin metal from ore. Approximately 50% of the metal processed in 1990 was recycled from scrap. (*Reynolds Aluminum Recycling Company*.)

of the aluminum beverage cans are recycled in the United States. See Figs. 2 and 3. Efforts also have been made to reduce the amount of aluminum needed for each bevarage can. The thickness of the aluminum can body has been reduced from an average weight of 21 grams/can in 1972 to approximately 15 grams/can in 1992. Can manufacturers have developed new designs that retain strength while reducing the amount of metal needed. In one innovation, the top of the can was "necked in"—first one, then two, three, and four times. Each successive necking down of the end lowered the weight of the can. By 1990, the average weight/can had been reduced by at least 26%.

Because of production process improvements, the cost of aluminum has been reduced by many fold over the past century. The demographics of aluminum production have changed markedly in recent years. Since 1982, the virgin metal producers in the United States have closed down or abandoned seven of the older, less-efficient, and higher-cost reduction plants. American producers have gone to other countries to build new modern plants using the most recently developed technology for large electrolytic cells. New generation electrolytic cells operate at very high current densities (180,000–275,000 amperes). These cells will produce four times the output of many existing cells while using 25% less energy.

The principal aluminum-producing areas of the world, in addition to the United States and Europe, are Canada, Venezuela, Brazil, Norway, Australia, and the Republic of the Congo. In the United States the remaining plants are located in the states of Washington, Oregon, and New York, along the Mississippi, and where water power generation is still available at somewhat reduced cost.

In 1960, 90% of world capacity was held by private enterprises. That percentage had dropped to 60% in 1985. The trend is to greater governmental participation in Latin America. Africa, the Mideast, and Asia. Fifty percent of Western World capacity is government sponsored, primarily in Europe and Latin America. Six corporate groups account for 50% of aluminum production. These are Alcan, Alcoa, Reynolds, Kaiser, Pechiney, and Alusuisse. They possess 40% of the bauxite capacity for alumina and aluminum. All are integrated from mine to ultimate consumer. Australia is the world's largest producer of bauxite.[1]

By 1995, Australia will probably be one of the world's three largest aluminum producers. At present, it supplies one-third of the world's bauxite. New smelter construction underway is expected to boost the world production by two million tons (1.8 million metric tons). New Zealand also has a good potential. Australia has coal reserves estimated at 1500 years.

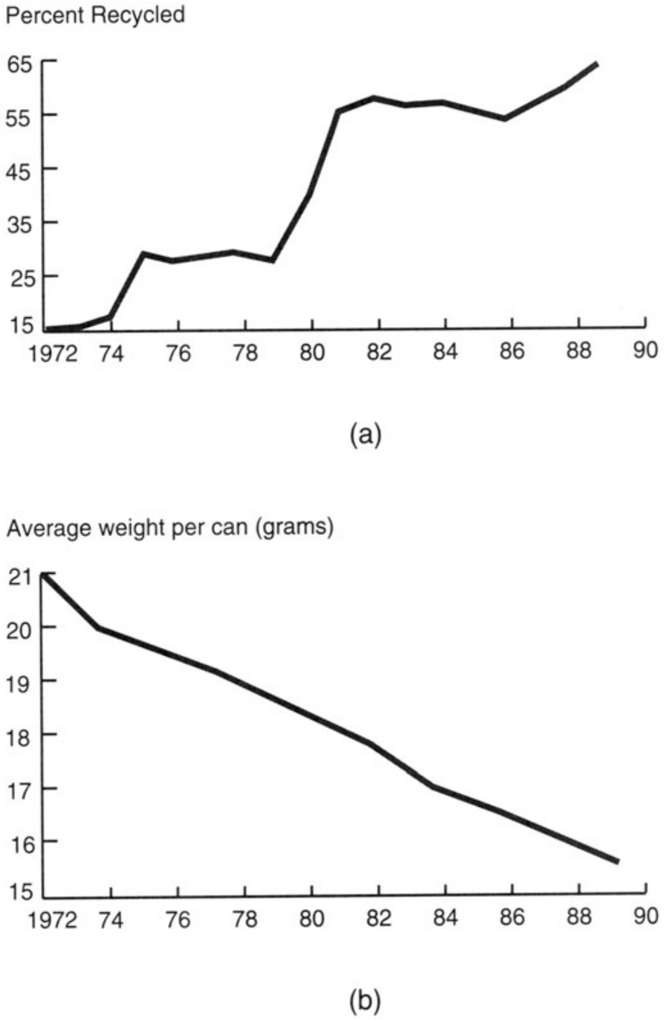


Fig. 3. Steps to conserve aluminum and the electrical power to produce it: (a) Percent of aluminum cans recycled (1972-1990); (b) average weight of beverage can (grams) (1972-1990). (Source: *Aluminum Association*, Washington, D.C.)

It is also projected that Venezuela, because of its abundance of hydropower, will become a major aluminum producer. During the past decade, Brazil has attracted considerable interest because of its vast reserves of good quality bauxite and hydropower. The Tucurui hydroelectric power facility will double its output by 1995.

The People's Republic of China is modernizing with the help and technology of foreign companies to bring its production to over 500,000 metric tons.

Most of the increased capacity for virgin aluminum production for the past two and a half decades has occurred outside the continental United States. The annual growth rate for the aluminum industry during this period is estimated at about 1.6%. The projected U.S. growth rate is about 1% per year, as contrasted with $2\frac{1}{4}\%$ worldwide. U.S. consumption is about 50 pounds (22.5 kg) per person per year. World consumption has doubled about every 20 years. The developed nations (North America, Western Europe, Japan-Oceania) account for two-thirds of world consumption of aluminum (1991).

Thermal Reduction Process. Because of the large electric power requirements of the electrolytic process, an effort has been underway for several years to cut production costs. The thermal process, including the use of electric arc furnaces, make it possible to achieve a greater concentration of electric energy in the refining process. Commercial electric furnace processes are fed with bauxite and clays (kaolin or kyanite or partially refined bauxite mixed with carbon or charcoal in briquetted form). An electric dc or ac arc is used. To date, this process has

---

[1] Firms participating in the development of Australian production capacity include Alcan, Alcoa, Kaiser, Comalco, Alumax, Reynolds, Pechiney and several Japanese companies.

been practical only for the production of impure aluminum alloys, ranging in content from 65–70% aluminum, 25–30% silicon, and 1% or more of iron. Other impurities may include titanium and the carbides and nitrides of aluminum and iron. This impure product may be refined to yield pure aluminum by means of a subsequent distillation process in which the electric-furnace product is dissolved at high temperatures in zinc, magnesium, or mercury baths and then distilled. Aluminum with a purity of 99.99+% may be produced in this manner.

A gas-reaction purification process also has attracted attention over the years. This is based on reacting aluminum trichloride gas with molten aluminum at about 1000°C to produce aluminum monochloride gas. Aluminum fluoride gas may be substituted for the aluminum monochloride gas in the process. Raw materials for the process may be scrap aluminum, aluminum from thermal-reduction processes, aluminum carbide, or aluminum nitride. As of the beginning of 1987, none of these other processes are fully commercial.

For the production of superpurity aluminum on a large scale, the Hoopes cell is used. This cell involves three layers of material. Impure (99.35 to 99.9% aluminum) metal from conventional electrolytic cells is alloyed with 33% copper (eutectic composition) which serves as the anode of the cell. A middle, fused-salt layer consists of 60% barium chloride and 40% $AlF_3 1.5NaF$ (chiolite), mp 720°C. This layer floats above the aluminum-copper alloy. The top layer consists of superpurity aluminum (99.995%). The final product usually is cast in graphite equipment because iron and other container metals readily dissolve in aluminum. For extreme-purity aluminum, zone refining is used. This process is similar to that used for the production of semiconductor chemicals and yields a product that is 99.9996% aluminum and is available in commerical quantities.

Continuing Research and Development. The industry continues to research improvements in the present production cells. Special attention is being focused on developing inert anodes and cathodes. Ferrites may find use as inert anodes, while titanium diboride may become the optimum material for cathodes. Before commercial use of inert electrodes can be achieved, cell sidewall materials must be developed which will withstand extremely reactive conditions and further improvements (i.e., less solubility of the anode and cathode materials are required). Over the past 15 years, American and Canadian aluminum producers have channeled nearly $1.5 billion into manufacturing technology research, the modernization and computerization of plant facilities, and new and better applications for the metal. Some of the results achieved thus far include:

- Energy efficiency—improved by almost 20%.
- New smelters with large, more efficient electrolytic cells.
- Recycling processes for all types of aluminum scrap, including aluminum beverage cans.
- Electromagnetic casting (EMC) for aluminum ingots in which an electromagnetic force field acts as the mold, thus producing smooth surfaces. In many cases, this eliminates wasteful scalping.
- Fluxing and degassing methods, combined with filtration, to give highest quality aircraft and other critical alloys.
- Continuous casting of filtered molten metal into giant coils (50,000 lb; 22,680 kg) of 0.250-inch (~6.4 mm) strip. This eliminates all hot rolling and intermediate annealing for rolling thin sheet products, such as foil, container stock, siding, etc.
- Aluminum particle rolling directly into continuous strip. Molten metal is sprayed into particles about the size of rice in a controlled environment, then brought to a desired rolling temperature and continuously rolled into "instant" coiled sheet This process can handle alloys which normally cannot be cast in ingot form.
- Computerized-automated cold rolling mills capable of 8000 ft (2438 m) per minute speeds. Multi-stand cold mills eliminate coil changes and wasted time. See Fig. 4.
- Foil rolling at speeds up to 8000 ft (2438 m) per minute with automatic gauge and shape control.
- Aluminum cans have captured over 90% of the beverage industry in the United States. The intent now is to capture a major share of the 30 billion cans food packaging market. (This will be challenged by the increasing use of plastics.)
- Increased use of aluminum in automotive vehicles. Approximately 200 pounds (90 kg) of aluminum per U.S.-built vehicle (i.e., wheels, engine blocks, bumpers, heat exchangers, and other parts). See Fig. 5.
- New alloys, such as aluminum-lithium alloys for aircraft applications which will permit 7 to 10% weight reduction per aircraft.

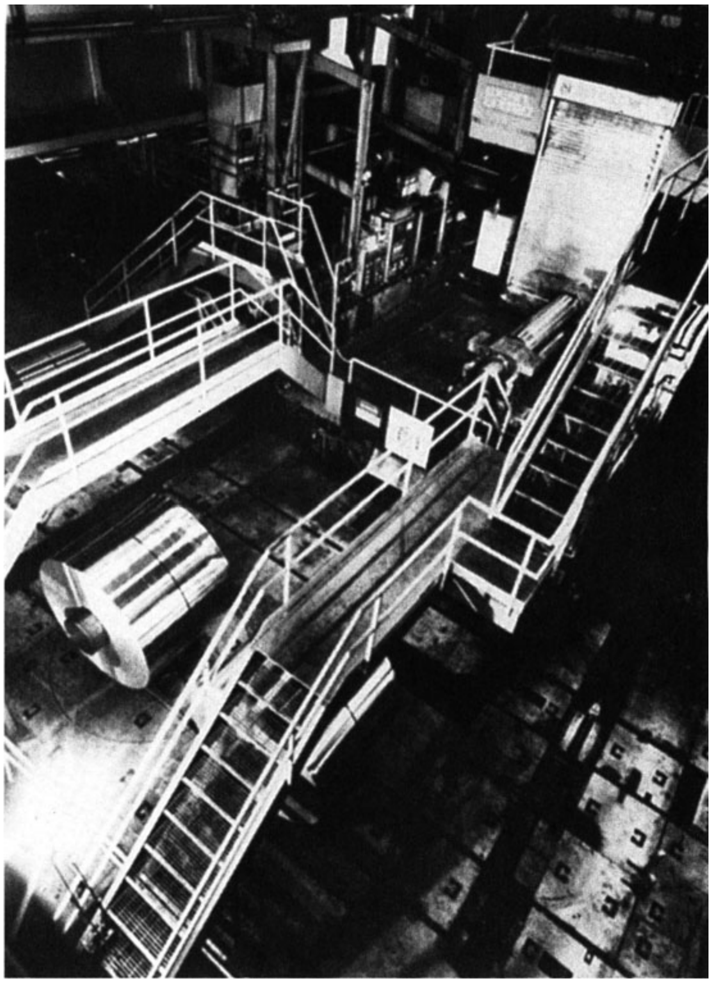


Fig. 4. High-speed, computer-controlled cold rolling mill can roll coils up to 40,000 pounds (18,150 kg), at speeds up to 5500 feet (1675 m) per minute. The system is equipped with automatic hydraulic gage control and shape control systems.(*Reynolds Metals Company*.)

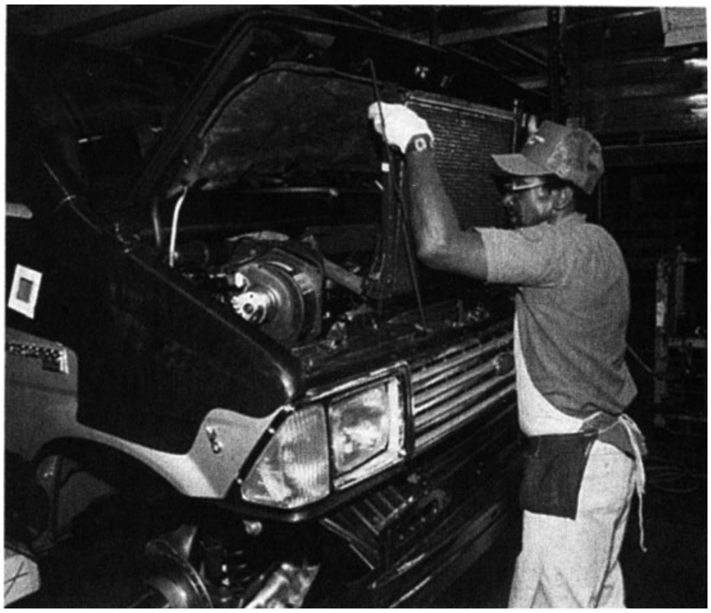


Fig. 5. All-aluminum radiator being installed in an automotive vehicle.

- Aluminum powder metallurgy continues to receive high research and development priority. A major objective is to produce final aluminum parts by direct pressing of *rapidly solidified aluminum particles*. The process handles high-strength aluminum alloy materials.
- Aluminum composites aim at combining high-strength aluminum alloys

with alumina ceramic fibers and filaments to provide even higher-strength structural materials.

- Superplastic—high-strength aluminum alloys which can be plastically shaped into difficult designed parts used on cars and aircraft.
- Rheocasting technology for producing stronger castings by combining casting and forging of molten aluminum alloys in one step.
- Ultrasonic welding, used increasingly in the aerospace, automotive, electrical, and electronics industries. The process requires simpler surface preparation and less energy and allows welding of thinner sheets to thicker structural members.
- Inertia (friction) welding in the solid state, which is useful for joining dissimilar metals, i.e., aluminum to steel and stainless steel. At least one of the pieces to be welded must be circular and capable of being rotated at high velocities before being brought into contact with pressure to the second member.
- Vacuum brazing for producing heat exchangers.
- Aluminum as energy (fuel) source. The high energy content of aluminum has made it a logical choice as a sold fuel for boosting vehicles, such as the space shuttle, into orbit.
- Aluminum—air battery. A second potential application of this available energy is based on electrochemical oxidation of aluminum in air to produce electricity. In an aluminum—air battery, for example, thin coils of aluminum strip may be used as the fuel. No electric battery recharging would be required since the aluminum is consumed to generate the electricity directly. This fuel would not give off fumes or pollute and could be stored in solid form indefinitely. If this concept materializes into commercial viability, it will provide the energy needed for electric vehicles.

Some of the foregoing topics are described in more detail in the following article on **Aluminum Alloys and Engineered Materials.**

Aluminum and Space. The year 1986 marked the hundredth anniversary of the Hall-Heroult electrolytic process for producing aluminum economically. Jules Verne, the 19th Century fiction writer, foretold the use of aluminum for space travel while it was still a mystery metal. Aluminum is the key metal in the aerospace industry. The Wright brothers recognized its value on their first flight in 1903 by using aluminum to build part of their engine. In military and commerical aircraft, aluminum became the metal of construction as early as 1930 for the U.S. Army P-12 and the Navy F-48, as well as an aluminum dirigible ZNC-2. World War II "light metals war" planes (i.e., B-17, B-25, B-29, etc.) became the forerunners of present aircraft. The U.S. Army Ballistic Missile Agency launched a program in 1950 for surface-to-surface missiles—the Redstone and Jupiter rocket programs. These programs prepared the way for NASA to launch unmanned satellites and to put men on the moon. Aluminum was the metal used for the rockets, the space modules, the moon rover (land vehicle), and most equipment associated with the programs. Today aluminum is used in the solid fuel booster rockets (some containing 0.5 mil lb—0.23 mil kg—of aluminum powder), in the space shuttles and planned space station. Some commercial airliners (jumbo jets 747 and DC-10) contain more than 300,000 lb (~136,000 kg) of aluminum each.

ALUMINUM CHEMISTRY AND CHEMICALS

Since aluminum has only three electrons in its valence shell, it tends to be an electron acceptor. Its strong tendency to form an octet is shown by the tetrahedral aluminum compounds involving sp^3 hybridization. Aluminum halides include the trifluoride, an ionic crystalline solid, the trichloride and tribromide both of which are dimeric in the vapor and in nonpolar solvents, having the halogen atoms arranged in tetrahedra about each aluminum atom, giving a bridge structure; and the triiodide. They combine readily with many other molecules, especially organic molecules having donor groups, and aluminum chloride is an active catalyst. Certain of the hydrates have a saltlike nature; thus $AlCl_3 \cdot 6H_2O$ acts as a salt, even though Al_2Cl_6 is covalent.

Common water-soluble salts include the sulfate, selenate, nitrate, and perchlorate, all of which are hydrolyzed in aqueous solution. Double salts are formed readily by the sulfate and by the chloride.

There are many fluorocomplexes of aluminum. The general formula for the fluoroaluminates is $M^1_x[Al_yF_{x+3y}]$, based upon AlF_6 octohedra, which may share corners to give other ratios of Al:F than 1:6. Chloroaluminates of the type $M^1[AlCl_4]$ are obtainable from fused melts. Aluminum ions form chloro-, bromo-, and iodo-complexes containing tetrahedral $[AlX_4]^-$ ions. However, in sodium aluminum fluoride $NaAlF_4$, the aluminum atoms are in the centers of octohedra of fluorine atoms in which the fluorine atoms are shared with neighboring aluminum atoms.

Aluminum hydroxide is about equally basic and acidic, the pK_b being about 12 and the pK_a being 12.6. Sodium aluminate seems to ionize as a uni-univalent electrolyte: $NaAl(OH)_4(H_2O)_2 \leftrightarrows Na^+ + [Al(OH)_4(H_2O)_2]^-$. The high viscosity of sodium aluminate solutions is explained by hydrogen bonding between these hydrated ions, and between them and water molecules.

By reaction of aluminum and its chloride or bromide at high temperature, there is evidence of the existence of monovalent aluminum. Here the aluminum atom is apparently in the *sp* state, with an electron pair on the side away from the chlorine atom, whereby the single pairs on the two chlorine atoms are shared to form two weak π bonds.

Aluminum forms a polymeric solid hydride, $(AlH_3)_x$, unstable above 100°C, extremely reactive, e.g., giving hydrogen explosively with H_2O and igniting explosively in air. From it are derived the tetrahydroaluminates, containing the ion AlH_4^-, which are important, powerful but selective reducing agents, e.g., reducing chlorophosphines R_2PCl to phosphines R_2PH; reactive alkyl halides and sulfonates to hydrocarbons, epoxides, and carboxylic acids; aldehydes and ketones to alcohols; nitrites, nitro compounds and N,N-dialkyl amides to amines. The salt most commonly used is lithium (tetra)hydroaluminate ("lithium aluminum hydride").

Aluminum compounds are generally made starting with bauxite, which is reactive with acids and with bases. With acids, e.g., H_2SO_4, any iron contained in the bauxite is dissolved along with the aluminum and silicon is left in the residue, whereas with bases, e.g., NaOH, any silicon is dissolved and iron left in the residue.

Acetate: Aluminum acetate $Al(C_2H_3O_2)_3$, white crystals, soluble, by reaction of aluminum hydroxide and acetic acid and then crystallizing. Used (1) as a mordant in dyeing and printing textiles, (2) in the manufacture of lakes, (3) for fireproofing fabrics, (4) for waterproofing cloth.

Aluminates: Sodium aluminate $NaAlO_2$, white solid, soluble, (1) by reaction of aluminum hydroxide and NaOH solution, (2) by fusion of aluminum oxide and sodium carbonate; the solution of sodium aluminate is reactive with CO_2 to form aluminum hydroxide. Used as a mordant in the textile industry, in the manufacture of artificial zeolites, and in the hardening of building stones. See silicates below and calcium aluminates.

Alundum: See oxide (below).

Carbide: Aluminum carbide Al_4C_3, yellowish-green solid, by reaction of aluminum oxide and carbon in the electric furnace, reacts with H_2O to yield methane gas and aluminum hydroxide.

Chlorides: Aluminum chloride $AlCl_3 \cdot 6H_2O$, white crystals, soluble, by reaction of aluminum hydroxide and HCl, and then crystallizing; anhydrous aluminum chloride $AlCl_3$, white powder, fumes in air, formed by reaction of dry aluminum oxide plus carbon heated with chlorine in a furnace, used as a reagent in petroleum refining and other organic reactions.

Fluoride: Aluminum fluoride AlF_3, white solid, soluble, by reaction of aluminum hydroxide plus hydrofluoric acid and then crystallizing $2AlF_3 \cdot 7H_2O$, used in glass and porcelain ware.

Hydroxide: Aluminum hydroxide $Al(OH)_3$, white gelatinous precipitate, by reaction of soluble aluminum salt solution and an alkali hydroxide, carbonate or sulfide (sodium aluminate is formed with excess NaOH but no reaction with excess NH_4OH), upon heating aluminum hydroxide the residue formed is aluminum oxide. Used as intermediate substance in transforming bauxite into pure aluminum oxide.

Nitrate: Aluminum nitrate $Al(NO_3)_3$, white crystals, soluble, by reaction of aluminum hydroxide and HNO_3, and then crystallizing.

Oleate: Aluminum oleate $Al(C_{18}H_{33}O_2)_3$, yellowish-white powder, by reaction of aluminum hydroxide, suspended in hot H_2O, shaken with oleic acid, and then drying, the product is used (1) as a thickener for lubricating oils, (2) as a drier for paints and varnishes, (3) in waterproofing textiles, paper, leather.

Oxide: Aluminum oxide, alumina Al_2O_3, white solid, insoluble, melting point 2020°C, formed by heating aluminum hydroxide to decomposition; when bauxite is fused in the electric furnace and then cooled, there results a very hard glass ("alundum"), used as an abrasive (hardness 9 Mohs scale) and heat refractory material. Aluminum oxide is the only oxide which reacts both in H_2O medium and at fusion temperature, to form salts with both acids and alkalis.

Palmitate: Aluminum palmitate $Al(C_{16}H_{31}O_3)_3$, yellowish-white powder, by reaction of aluminum hydroxide, suspended in hot H_2O, shaken with palmitic acid, and then drying, the product is used (1) as a thickener for lubricating oils, (2) as a drier for paints and varnishes, (3) in waterproofing textiles, paper, leather, (4) as a gloss for paper.

Silicates: Many complex aluminosilicates or silicoaluminates are found in nature. Of these, clay in more or less pure form, pure clay, kaolinite, kaolin, china clay $H_4Si_2Al_2O_9$ or $Al_2O_3{\cdot}2SiO_2{\cdot}2H_2O$ is of great importance. Clay is formed by the weathering of igneous rocks, and is used in the manufacture of bricks, pottery, procelain, and Portland cement.

Stearate: Aluminum stearate $Al(C_{18}H_{35}O_2)_3$, yellowish-white powder, by reaction of aluminum hydroxide suspended in hot H_2O, shaken with stearic acid, and then drying the product, used (1) as a thickener for lubricating oils, (2) as a drier for paints and varnishes, (3) in waterproofing textiles, paper, leather, (4) as a gloss for paper.

Sulfate: Aluminum sulfate $Al_2(SO_4)_3$, white solid, soluble, by reaction of aluminum hydroxide and H_2SO_4, and then crystallizing, used (1) as a clarifying agent in water purification, (2) in baking powders, (3) as a mordant in dyeing, (4) in sizing paper, (5) as a precipitating agent in sewage disposal; aluminum potassium sulfate, see **Alum** and **Alunite**.

Sulfide: Aluminum sulfide Al_2S_3, white to grayish-black solid, reactive with water to form aluminum hydroxide and H_2S, formed by heating aluminum powder and sulfur to a high temperature.

Aluminum in solution of its salts is detected by the reaction (1) with ammonium salt of aurin tricarboxylic acid ("aluminon"), which yields a red precipitate persisting in NH_4OH solution, (2) with alizarin red S, which yields a bright red precipitate persisting in acetic acid solution.

Organoaluminum Compounds: By the action of alpha olefins on AlH_3, the trialkyls of aluminum can be prepared:

$$AlH_3 + 3CH_2{=}CHR \xrightarrow{120°C} (CH_2CH_2R)_3Al$$

In the presence of ethers, magnesium-alloy alloys react with alkyl halides to yield trialkyls, R_3AlOEt_2. The trimethyl aluminum compound is a dimer, mp 15°C. Triethyl aluminum is a liquid. The alkyls react readily with H_2O:

$$Et_3Al + 3H_2O \rightarrow Al(OH)_3 + 3EtH$$

When aluminum is reacted with diphenyl mercury $(C_6H_5)_2Hg$, triphenylaluminum $(C_6H_5)_3Al$ is yielded. The aluminum organometallic alkyls and aryls act as Lewis acids to form compounds with electron-donating substances. The resulting compounds are the organometallic basis for producing polymers of the Al-N type.

Zeolite Structures: These are crystalline, microporous solids that contain cavities and channels of molecular dimensions (3Å to 10Å) and sometimes are called *molecular sieves*. Zeolites are used principally in catalysis, separation, purification, and ion exchange. The fundamental building block of a zeolite is a tetrahedron of four oxygen atoms surrounding a central silicon atom (i.e., $(SiO_4)^{4-}$). From the fundamental unit, numerous combinations of secondary building units (polygons) can be formed. The corners of these polyhedra may be Si or Al atoms[2]. Different combinations of the same secondary building unit may yield numerous distinctive zeolites. As pointed out by D. E. W. Vaughan, few fields of chemistry offer such chemical and structural diversity. Although only about sixty structures are known, tens of thousands of theoretical structures are possible. The science of synthesizing zeolites, which achieves much more diversity than is the case with natural zeolites, has been under continuous development for several years. Synthetic zeolites often are cataloged by their Si/Al composition ratios, usually ranging from 1:1 to 10:500. Thus, aluminum plays a large role in this field. In the manufacture of synthetic zeolites, the common reactants are silica from sodium silicate (water-glass) or an aqueous colloidal silica, aluminum from sodium aluminate, alum, or a colloidal alumina (boehmite), and sodium or potassium hydroxide. For further detail, see the Vaughan (1988) reference.

Additional Reading

Note: See list of references at end of next article.

S. J. Sansonetti, Consultant, Reynolds Metals Company,
Richmond Virginia. (Updated by editorial staff.)

ALUMINUM ALLOYS AND ENGINEERED MATERIALS. Aluminum alloys have been used effectively for scores of years. Much research during the past few decades has been directed not only to expand the list of alloys, but also for applying aluminum in other engineered forms, such as composites, sandwiched, and laminated materials. Aluminum fabrication and surface treatment have been improved. The techniques of powder metallurgy have been expanded to include the new engineering forms of aluminum. Aluminum quasicrystals are a recent discovery. The "shape memory" of certain aluminum-base materials is being exploited. A process for using shock waves to shape aluminum also is among recent achievements.

Anually, sessions are held on *advanced aluminum alloys* as part of the conference on "Advanced Aerospace Materials/Processess and Exposition." Topics discussed have included rapid-solidification alloys, very low density alloys, laminates, new alloying elements, such as lithium, and such processes as diffusion bonding and means to improve fracture toughness and reduce fatigue and corrosion.

Aluminum is unique among the metals because it responds to nearly all of the known finishing processes. It can be finished in the softest, most delicate textures as exemplified by tableware and jewelry. Aluminum can be anodized and dyed to appear like gold. It can be made as specular as a silver mirror and jet black. The metal also can be anodized to an extremely hard, wear- and abrasion-resistant surface that approaches the hardness of a diamond. Aluminum is available in many convenient forms—shapes, sheet, plate, ingot, wire, rod and bar, foil, castings, forgings, powdered metals, and extrusions.

In the production of these products, new high-speed automated equipment has been developed over the past decade. Rolling speeds as high as 8000 feet (2438 meters) per minute are possible on some products. New continuous casting processes have been developed where molten metal can be converted directly into coiled sheet, using roll casting, belt casting, and articulated block casting. See Figs. 1 and 2. Many of the common alloys respond well to these new processes. Electromagnetic casting of the conventional giant ingots used for the conventional rolling mills has been developed in the fully automated mode. This casting process eliminates the need for scalping ingots prior to processing into final products, such as can stock, foil, auto body sheet, siding, and many other products. In this process, an electrical field rather than a solid mold wall is used to shape the ingot. The process is based upon the fact that liquids with good electrical conductivity, such as aluminum, can be constrained where an alternating magnetic field is applied. The liquid metal remains suspended by the electrical forces while it is being solidified.

The casting alloy products comprise sand castings, permanent mold castings, and die castings. Aluminum is the basic raw material for more than 20,000 businesses in the United States. Aluminum is an indispensable metal for aircraft, for example. See Fig. 3. Representative aluminum alloys for a broad classification of uses is given in Table 1.

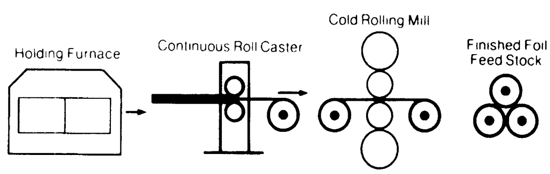


Fig. 1. Schematic portrayal of continuous aluminum roll casting process. (*Reynolds Metals Company*.)

---

[2] Other atoms are used, but to a lesser extent, including Li, Be, B, Mg, Co, Mn, Zn, P, As, Ti, Ga, Fe, and Fe.

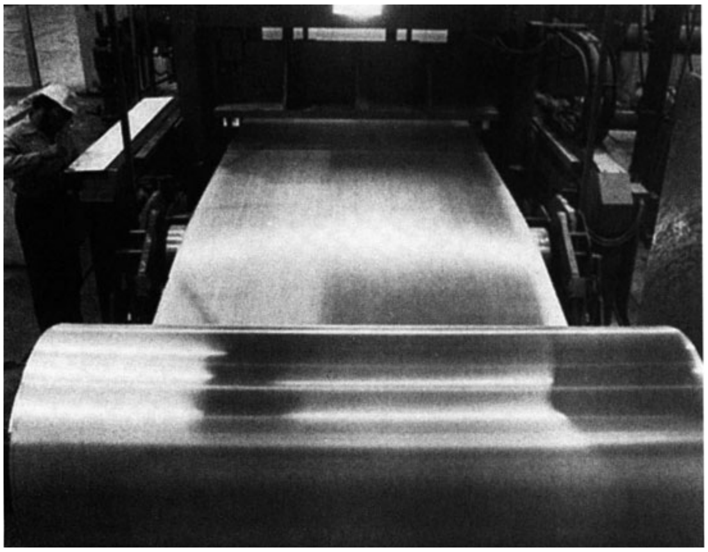


Fig. 2. Continuous rolling plant (Hot Spring, Arkansas) has cast nearly two billion pounds (0.9 billion kg) of aluminum foil feedstock since it began operations in 1979. That quantity of foil would wrap the earth almost 1600 times at the equator. The plant produces coils of feedstock for aluminum foil and flexible packaging products at a 40% energy saving compared to conventional rolling processes. (*Reynolds Metals Company.*)

Fig. 3. Final operations on machined aluminum plates for wings of large jet aircraft. (*Reynolds Metals Company.*)

Worldwide demand for aluminum over the past several decades has increased by approximately 5–6% per year. Aluminum is quite well established as an energy saver and a cost-effective material in numerous applications. Particularly strong growth is occurring in the automotive and transportation fields, housing, and food packaging. For example, a number of breakthroughs in foil pouch packaging have been achieved for replacing conventional cans for shelf-stable foods. Steady growth has continued in other areas, including aircraft and space applications, electrical wire and cable, marine, architectural, and consumer durable products. See Fig. 4.

Pure aluminum is soft and ductile. In its annealed form, the tensile strength is approximately 13,000 pounds per square inch (90 megapascals). This can be work hardened by rolling, drawing into wire, or by other cold-working techniques to increase its strength to approximately 24,000 pounds per square inch (165 megapascals). Pure aluminum is particularly useful in the food and chemical industries where its resistance to corrosion and high thermal conductivity are desirable characteristics, and in the electrical industry, where its electrical conductivity of about 62% that of copper and its lightweight make it desirable for wire, cable, and bus bars. Large quantities are rolled to thin foils for the packaging of food products, for collapsible tubes, and for paste and powders used for inks and paints.

Most commercial uses require greater strength than pure alumnium affords. Higher strength is achieved by alloying other metal elements with aluminum and through processing and thermal treatments. The alloys can be classified into two categories, non-heat-treatable and heat-treatable.

Non-heat-Treatable Alloys. The initial strength of alloys in this group depend upon the hardening effect of elements such as manganese, silicon, iron, and magnesium, singly or in various combinations (see three alloys in Table 2). The non-heat-treatable alloys are usually designated therefore in the 1xxx, 3xxx, 4xxx, and 5xxx series (Table 3). Since these alloys are work hardenable, further strengthening is made possible by various degrees of cold work denoted by the H series tempers.

Heat-Treatable Alloys. This group of alloys includes the alloying elements, copper, magnesium, zinc, and silicon (-T tempered alloys in Table 2). Since these elements singly or in combination show increasing solid solubility in aluminum with increasing temperature, it is possible to subject them to thermal treatments which will impart pronounced strengthening. The first step called heat treatment or solution heat treatment, is an elevated temperature process designed to put the soluble element or elements in solid solution. This is followed by a rapid quenching which monetarily "freezes" the structure and for a short time renders the alloy very workable. At room or elevated temperatures, the alloys are not stable after quenching, and precipitation of the constituent from the supersaturated solid solution begins. After several days at room temperature, termed aging or room temperature precipitation, the alloy is considerably stronger. Many alloys approach a stable condition at room temperature but some alloys, especially those containing magnesium and silicon or magnesium and zinc continue to age harden for long periods of time at room temperature. By heating for a controlled time at slightly elevated temperatures, even further strengthening is possible and the properties are stabilized.

Clad Alloys. The heat-treatable alloys, in which copper or zinc are major alloying constituents, are less resistant to corrosive attack than a majority of the non-heat-treatable alloys. To increase the corrosion resistance of these alloys in sheet and plate form, they are often clad with a high-purity aluminum, a low-magnesium-silicon alloy, or an aluminum alloy containing 1% zinc.

The cladding, usually from $2\frac{1}{2}$ to 5% of the total thickness on each side, not only protects the composite due to its own inherent excellent corrosion resistance, but also exerts a galvanic effect which further protects the core alloy.

Special composites may be obtained such as clad non-heat-treatable alloys for extra corrosion protection, for brazing purposes, and for special surface finishes.

Composites. The most commonly used metal elements alloyed with aluminum are magnesium, manganese, silicon, copper, zinc, iron, nickel, chromium, titanium, and zirconium. The strength of aluminum can be tailored to the specific end applications ranging from the very soft ductile foils to the high-strength aircraft and space alloys equal to steels in the 90,000 pounds per square inch (621 megapascals) tensile strength range. In addition to the homogeneous aluminum alloys, dispersion hardened and advanced filament-aluminum composites provide even higher strengths. Boron filaments in aluminum can provide strengths in the range of 150,000–300,000 pounds per square inch (1034–2068 megapascals) tensile strength.

Superplasticity. Eutectoid and near eutectoid alloy chemistry research has resulted in alloys exhibiting superplasticity with unusual elongation (> 1000%) and formability and with 60,000 pounds per square inch (414 megapascals) tensile strength. Three examples of these alloys are: 94.5% A1-5% Cu-0.5% Zr; 22% A1-78% Zn; and 90% A1, 5% Zn, 5% Ca.

Casting Alloys. During the last two decades, the quality of castings has been improved substantially by the development of new alloys and better liquid-metal treatment and also by improved casting techniques. Casting techniques include sand casting, permanent mold casting, pressure die casting, and others. Today sand castings can be produced in high-strength alloys and are weldable. Die casting per-

TABLE 1. REPRESENTATIVE ALUMINUM ALLOYS

Product	Alloy	Form	Product	Alloy	Form
Architectural and Building Products					
Awnings	3003	Sheet			
Fence Wire	6061	Wire	Cable	E.C., 5005, 6201	Wire
Fittings	A514.0	Castings	Conduit	6063	Tubing
Gutters	Alc. 3004	Sheet	Motors	319.0, 355.0, 360.0, 380.0	Castings
Nails	5056, 6061	Wire			
Panels	3003, 6063	Sheet and Extrusions	Transmission Towers and Substations	6061, 6063, 7005	Extrusions
Roofing	Alc. 3004	Sheet			
Screens	Alc. 5056	Wire	*Consumer Durables*		
Siding	3003, Alc. 3004	Sheet	Appliances	3003, 4343, 5052, 5252, 5357, 5457, 6063, 6463	Sheet and Extrusions
Transportation					
Aircraft	2014, 2024, 2048, 2090, 2124, 2219, 7075, 7079, 7175, 7178, 7179, 7475, 8090, 8091, 8092.	Sheet, Plate, Extrusions, and Forgings		356.0, 390.0	Castings
	B295.0, 355.0, 356.0, 518.0, 520.0	Castings	Cooking Utensils	1100, 3003, Alc. 3003, 3004, Alc. 3004, 5052, 5454	Sheet
				B443.0, A514.0	Castings
			Furniture	3003, 3004, 5005, 6463	Sheet and Tubing
Automotive					
Auto and Truck Bodies	2036, 2037, 5052, 5182, 5252, 5657, 6009, 6010, 6061, 6063, 6463, 7016, 7021, 7029, 7046, 7129	Sheet and Extrusions	Refrigerators	3003, Alc. 3003, 3004, Alc. 3004, 4343, 5005, 5050, 5052, 5252, 5457, 6061, 6463	Sheet and Extrusions
Buses	2036, 5083, 5086, 5182, 5252, 5457, 5657, 6061, 6063, 6463, 7016, 7046	Sheet and Extrusions	Water Heaters	Alc. 6061, 3003	Sheet
			Machinery and Equipment		
	242.0, B295.0, 355.0, 356.0, 360.0, 380.0, 390.0, B443.0	Castings	Chemical Processing	1060, 1100, 3003, Alc. 3003, 3004, Alc. 3004, 5083, 5086, 5154, 5454, 6061, 6063	Sheet, Plate, Extrusions, and Tubes
				356.0, 360.0, B514.0	Castings
			Heat exchangers and Solar Panels	3003, Alc. 3003, 3004, Alc. 3004, 4343, 5005, 5052, 6061, 6951	Sheet and Tubing
Marine					
Barges, Small Craft, Ships, Tankers	5052, 5083, 5086, 5454, 5456, 6061, 6063	Sheet, Plate, Extrusions, and Tubing	Sheet for Vacuum Brazing Heat Exchangers	4003, 4004, 4005, 4044, 4104	
	360.0, 413.0 B443.0	Castings	Irrigation	3003, Alc. 3003, 3004, Alc. 3004, 5052, 6061, 6063	Tubing
Railroad	Alc. 2024, 5052, 5083, 5086, 5454 5456, 6061, 6063, 7005	Sheet, Plate, and Extrusions	Sewage Plants	3003, Alc. 3003, 3004, Alc. 3004, 5052, 5083, 5086, 5454, 5456, 6061, 6063	Sheet, Extrusions, and Tubes
	B295.0, 356.0, 520.0	Castings			
Containers and Packaging				B514.0	Castings
Foils	1100, 1235, 3003, 5005, 8079, 8111	Foils	Screw Machine Parts	2011, 2024, 6262	Rod and Bar
Cans	3003, 3004, 5182	Sheet	Textile Machinery	2014, 2024, 6061, 6063	Extrusions, Sheet and Plate
Electrical					
Bus Bar	6063, 6201	Extrusions or Rolled Rod	*Finished Pressed Parts*	7090, 7091	Powder Metallurgy

mits large production outputs per hour on intricate pieces that can be cast to close dimensional tolerance and have excellent surface finishes; hence, require minimum machinings. Since aluminum is so simple to melt and cast, a large number of foundry shops have been established to supply the many end products made by this method of fabrication. See Table 3.

Casting Semi-solid Metal. A new casting technology is based on vigorously agitating the molten metal during solidification. A very different metal structure results when this metal is cast. The vigorously agitated liquid-solid mixture behaves as a slurry still sufficiently fluid (thixotropic) to be shaped by casting. The shaping of these metal slurries is termed "Rheocasting."

The slurry nature of "Rheocast" metal permits addition and retention of particulate nonmetal (e.g., Al_2O_3, SiC, T, C, glass beads) materials for cast composites.

The importance of this new technology is beginning to be commercialized.

Alloy and Temper Designation Systems for Aluminum. The aluminum industry has standardized the designation systems for wrought aluminum alloys, casting alloys and the temper designations applicable. A system of four-digit numerical designations is used to identify wrought aluminum alloys. The first digit indicates the alloy group as shown in Table 4. The 1xxx series is for minimum aluminum purities of 99.00% and greater; the last two of the four digits indicate the minimum aluminum percentage; i.e., 1045 represents 99.45% minimum aluminum, 1100 represents 99.00% minimum aluminum. The 2xxx through 8xxx series group aluminum alloys by major allowing elements. In these series the first digit represents the major alloying element, the second digit indicates alloy modification, while the third and fourth serve only to identify the different alloys in the group. Experimental alloys are prefixed with an X. The prefix is dropped when the alloy is no longer considered experimental.

Cast Aluminum Alloy Designation System. A four-digit number system is used for identifying aluminum alloys used for castings and foundry ingot (see Table 5). In the 1xx.x group for aluminum purities of 99.00% or greater, the second and third digit indicate the minimum aluminum percentage. The last digit to the right of the decimal point indicates the product form: 1xx.0 indicates castings and 1xx.1 indicates ingot. Special control of one or more individual elements other than aluminum is indicated by a serial letter before the numerical designa-

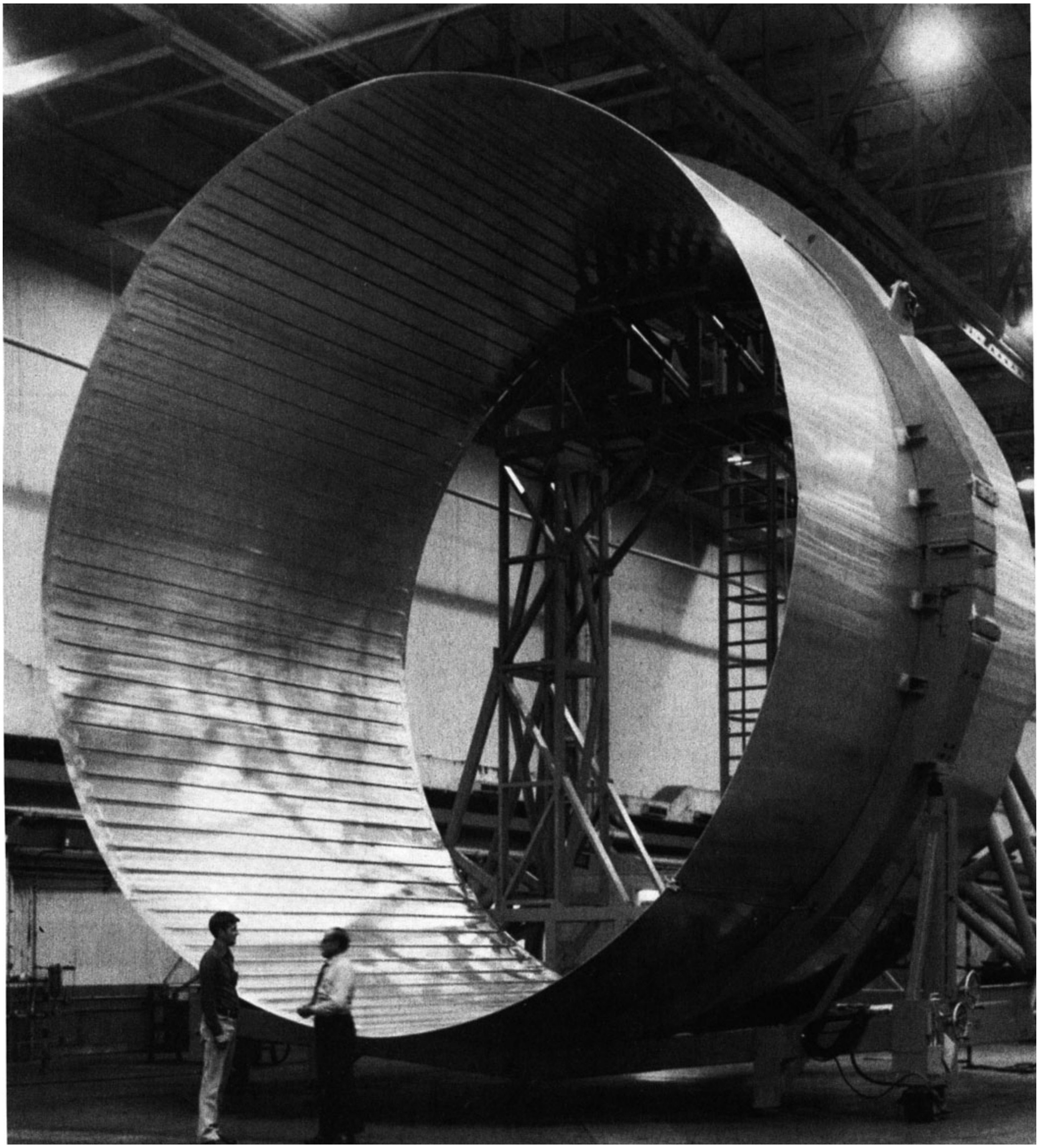


Fig. 4. Four aluminum sections of this type make up fuel tank for aerospace vehicle. Each section is produced from computer-machined plate and weighs 4750 pounds (2155 kg). (*Reynolds Metals Company.*)

tion. The serial letters are assigned in alphabetical sequence starting with A but omitting I, O, Q, and X, the X being reserved for experimental alloys.

In the 2xx.x through 9xx.x alloy groups, the second two of the four digits in the designation have no special significance but serve only to identify the different aluminum alloys in the group. The last digit to the right of the decimal point indicates the product form: .0 indicates casting and .1 indicates ingot. Examples: Alloy 213.0 represents a casting of an aluminum alloy whose major alloying element is copper. Alloy C355.1 represents the third modification of the chemistry of an aluminum alloy ingot whose major alloying elements are silicon, copper, and magnesium.

Temper Designation System. A temper designation is used for all forms of wrought and cast aluminum alloys. The temper designation follows the alloy designation, the two letters being separated by a hyphen. Basic designations consist of letters followed by one or more digits. These designate specific sequences of basic treatments but only operations recognized as significantly influencing the characteristics of the product. Basic tempers are -F (as fabricated), -O annealed (wrought products only), -H strain-hardened (degree of hardness is normally quarter hard, half hard, three-quarters hard, and hard designated by the symbols H12, H14, H16, and H18, respectively). -W solution heat treated and -T thermally treated to produce stable tempers. Examples: 1100-H14 represents commercially pure aluminum cold rolled to half-

TABLE 2. NOMINAL CHEMICAL COMPOSITION[1] AND TYPICAL PROPERTIES OF SOME COMMON ALUMINUM WROUGHT ALLOYS

	1100	3003	5052	2014T6[10]	2017T4[9]	2024T4[9]	6061T6[10]	7075T6[10]	6101T6[11]
Nominal chemical composition[1]	99% min. Alum.	1.2% Mn	2.5% Mg 0.25% Cr	4.4% Cu 0.8% Si 0.8% Mn 0.4% Mg	4.0% Cu 0.5% Mn 0.5% Mg	4.5% Cu 1.5% Mg 0.6% Mn	1.0% Mg 0.6% Si 0.25% Cu 0.25% Cr	5.5% Zn 2.5% Mg 1.5% Cu 0.3% Cr	0.5% Mg 0.5% Si
Tensile strength, psi	A 13,000[7]	16,000	28,000	—	—	—	—	—	—
	H 24,000[7]	29,000	42,000	70,000	62,000	68,000	45,000	83,000	32,000
Tensile strength, MPa	A 90	110	193	—	—	—	—	—	—
	H 165	200	290	483	427	469	310	572	221
Yield strength, psi[2]	A 5000	6000	13,000	—	—	—	—	—	—
	H 22,000	27,000	7,000	60,000	40,000	47,000	40,000	73,000	28,000
Yield strength, MPa	A 34	41	90	—	—	—	—	—	—
	H 152	186	255	414	276	324	276	503	193
Elongation percent in 2 in. (5.1 cm)[11]	A 45	A 40	A 30	13	22	19	17	11	15
	H 15	H 10	H 8						
Modulus of elasticity[3]	10	10	10.2	10.6	10.5	10.6	10	10.4	10
Brinnell hardness[8]	23–44	28–55	45–85	135	105	120	95	150	71
Melting range, °C	643–657	643–654	593–649	510–638	513–640	502–638	582–652	477–638	616–651
Melting range, °F	1190–1215	1190–1210	1100–1200	950–1180	955–1185	935–1180	1080–1250	890–1180	1140–1205
Specific gravity	2.71	2.73	2.68	2.80	2.79	2.77	2.70	2.80	2.70
Electrical resistivity[4]	2.9	3.4	4.93	4.31	5.75	5.75	4.31	5.74	3.1
Thermal conductivity[5]	0.53	A 0.46	A 0.33	0.37	0.29	0.29	0.37	0.29	0.52
SI units	221.9	A 192.6	A 138.2	154.9	121.4	121.4	154.9	121.4	217.7
Coefficient of expansion[6]	23.6	23.2	23.8	22.5	23.6	22.8	23.4	23.2	23

[1] Aluminum plus normal inpurities is the remainder.
[2] 0.2% permanent set.
[3] Multiply by 10^6.
[4] Microhms per cm (room temperature).
[5] C.g.s. units (at 100°C).
[6] Per °C (20–100°C); multiply by 10^{-6}
[7] A = annealed; H = hard.
[8] 500 kg load, 10 mm ball.
[9] Solution heat-treated and naturally aged.
[10] Solution heat-treated and artifically aged.
[11] Round specimens, $\frac{1}{2}$-m diameter.
Conversion factors used: 1 psi = 6.894757×10^{-3} megapascals (MPa)
C.g.s. = (cal)(cm^2)/(sec)(cm)(°C)
SI unit = Watts/meter°K
1 c.g.s. unit = 418.68 SI units

TABLE 3. NOMINAL CHEMICAL COMPOSITION[1] AND TYPICAL PROPERTIES OF SOME ALUMINUM CASTING ALLOYS
Properties for alloys 195, B195, 220, 355, and 356 are for the commonly used heat treatment.

	413.0[2]	B443.0[2]	208.0[3]	308.0[4]	295.0[3]	B295.0[4]	514.0[3]	518.0[2]	520.0[3]	355.0[3]	356.0[3]	380.0[2]
Nominal chemical composition	12% Si	5% Si	4% Cu	5.5% Si 3% Si	4.5% Cu 4.5% Cu	4.5% Cu 0.8% Si	3.8% Mg 2.5% Si	8% Mg	10% Mg	5% Si	7% Si 0.5% Mg	8.5% Si
Tensile strength, psi[5]	37,000	19,000	21,000	28,000	36,000	45,000	25,000	42,000	46,000	35,000	33,000	45,000
Tensile strength, MPa[5]	255	131	145	193	248	310	172	290	317	241	228	310
Yield strength, psi[5]	18,000	9000	14,000	16,000	24,000	33,000	12,000	23,000	25,000	25,000	24,000	25,000
Yield strength, MPa[5]	124	62	97	110	165	228	83	159	174	174	165	174
Elongation, percent[5]	1.8	6	2.5	2	5	5	9	7	14	2.5	4	2
Brinnell hardness[6]	—	40	55	70	75	90	50	—	75	80	70	—
Melting range, °C	574–585	577–630	521–632	—	549–646	527–627	580–640	540–621	449–621	580–627	580–610	521–588
Melting range, °F	1065–1085	1070–1165	970–1170	—	1020–1195	980–1160	1075–1185	1005–1150	840–1150	1075–1160	1075–1130	970–1090
Specific gravity	2.66	2.69	2.79	2.79	2.81	2.78	2.65	2.53	2.58	2.70	2.68	2.76
Electrical resistivity	4.40	4.66	5.56	4.66	4.66	3.45	4.93	7.10	8.22	4.79	4.42	6.50
Thermal conductivity[7]	0.37	0.35	0.29	0.34	0.35	0.45	0.33	0.24	0.21	0.34	0.36	0.26
SI units	154.9	146.5	121.4	142.4	146.5	188.4	138.2	100.5	87.9	142.4	150.7	108.9
Coefficient of expansion[8]	20.0	22.8	22.8	22.7	23.9	22.8	24.8	24.0	25.4	22.8	22.8	20.0

[1] Remainder is aluminum plus minor impurities.
[2] Die cast.
[3] Sand cast.
[4] Permanent mold cast.
[5] For separately cast test bars.
[6] 500 kg/load, 10 mm, ball.
[7] C.g.s. units.
[8] Multiply by 10^{-6}. Per °C, for temperature range 20 to 200°C.
Conversion factors used: 1 psi = 6.894757×10^{-3} megapascals (MPa)
SI unit = Watts/meter °K
1 c.g.s. unit = 418.68 SI units

TABLE 4. DESIGNATIONS FOR WROUGHT ALUMINUM ALLOY GROUPS

		Alloy No.
Aluminum—99.00% minimum and greater		1xxx
	Major Alloying Element	
Aluminum Alloys grouped by Major Alloying Elements	Copper	2xxx
	Manganese	3xxx
	Silicon	4xxx
	Magnesium	5xxx
	Magnesium and Silicon	6xxx
	Zinc	7xxx
	Other Element	8xxx
Unused Series		9xxx

(1) For codification purposes an alloying element is any element which is intentionally added for any purpose other than grain refinement and for which minimum and maximum limits are specified.

(2) Standard limits for alloying elements and impurities are expressed to the following places:

Less than 1/1000%	0.000X
1/1000 up to 1/100%	0.00X
1/100 up to 1/10%	
Unalloyed aluminum made by a refining process	0.0XX
Alloys and unalloyed aluminum not made by a refining process	0.0X
1/10 through 1/2%	0.XX
Over 1/2%	0.X, X.X, etc.

TABLE 5. DESIGNATIONS FOR CAST ALUMINUM ALLOY GROUPS

		Alloy No.
Aluminum	99.00% minimum and greater	1xx.x
	Major Alloy Element	
	Copper	2xx.x
Aluminum Alloys grouped by Major Alloying Elements	Silicon, with added Copper and/or Magnesium	3xx.x
	Silicon	4xx.x
	Magnesium	5xx.x
	Zinc	7xx.x
	Tin	8xx.x
	Other Element	9xx.x
Unused Series		6xx.x

(1) For codification purposes an alloying element is any element which is intentionally added for any purpose other than grain refinement and for which minimum and maximum limits are specified.

(2) Standard limits for alloying elements and impurities are expressed to the following places:

Less than 1/1000%	0.000X
1/1000 up to 1/100%	0.00X
1/100 up to 1/10%	
Unalloyed aluminum made by a refining process	0.0XX
Alloys and unalloyed aluminum not made by a refining process	0.0X
1/10 through 1/2%	0.XX
Over 1/2%	0.X, X.X, etc.

hard properties. 2024-T6 represents an aluminum alloy whose principal major element is copper which has been solution heat treated and then artificially aged to develop stable full-strength properties of the alloy.

S. J. Sansonetti, Consultant, Reynolds Metals Company,
Richmond, Virginia.

Contemporary Advancements and Future Potential

Highlighted in the following paragraphs are improvements in aluminum metallurgy that have occurred and have been available only relatively recently or that are promising but that still remain in a late phase of research or testing.

Aluminum-Lithium Alloys. Both private and government funding have been invested in Al-Li alloy research for several years. As of the early 1990s, exceptionally good results have been been achieved by way of increasing the strenth-to-weight ratio and the stiffness of Al-Li alloys. Low ductility in the short-transverse direction has been a difficult problem to solve. Wide usage awaits further problem solving and testing for critical applications. The Al-Li alloy 2091-T3 (Pechiney) is a medium-strength, lighweight alloy quite similar to the traditional alloy 2024-T3, which it is expected to replace for aerospace applications. The new alloy has a 7% lower density and a 10% higher stiffness. The new alloy, like most Al alloys, is notch sensitive. An oxide film composed of MgO, LiO_2, $LiAlO_2$, Li_2CO_3, and LiOH tends to develop under normal production conditions. Cracks form in this film and tend to initiate cracks in the alloy's substrate and this reduces fatigue life. When the film is removed, in both longitudinal and long-transverse directions, the new alloy's fatigue properties are comparable with other aluminum alloys.

In late 1989, the availability of a proprietary family of weldable, high-strength (*Weldalite*) Al-Li products appeared. The material was claimed to be nearly twice as strong (100×10^3 psi) as other leading alloys then currently used for aerospace applications. The alloy was initially developed especially for space-launch systems. Specific advantages claimed include: (1) high strength over a broad temperature range, from cryogenic to highly elevated temperatures, (2) light weight, and (3) weldability—this property being of particular value for fabricating fuel and oxidizer tanks for space vehicles. *Weldalite* is produced in sheet, plate, extrusion, and ingot products.

Al-Li investment castings are gaining acceptance. Among aluminum alloying elements, lithium is one of the most soluble. About 4.2% Li can be dissolved in Al at the eutectic temperature, 1116°F (602°C). However, in commercial-size ingots, the maximum Li content that can be cast without cracking is about 2.7%. Lithium is a strengthening element because of the formation of small, coherent ordered Al_3Li precipitates during aging (secondary hardening when Li content exceeds 1.4%). The toughness of Al-Li alloys, unlike conventional Al alloys, does not increase with increasing aging temperature (beyond that point needed for peak strength).

Metal-Matrix Composites. Silicon carbide particles are contributing to easy-to-cast metal-matrix composites (MMCs). When compared with their unreinforced counterparts, the SiC_p/Al components are more wear resistant, stiffer, and stronger, accompanied by improved thermal stability. Additional advantages include lower density and lower cost. Nearly all prior aluminum MMCs required labor-intensive methods, such as powder metallurgy, diffusion bonding, squeeze casting, or thermal spraying.

The new SiC composites are available as foundry ingot or extrusion billets. A new process ensures complete wetting of the SiC particles by molten aluminum. A number of investment castings are now being made, including aircraft hydraulic components and other small parts. These composites have excellent prospects for use in a variety of small parts, including medical prostheses and golf club heads.

Sialons consist of three-dimensional arrays of (Si-Al) $(O,N)_4$ tetrahedra. These oxynitrides are traditionally fabricated with silicon nitride. An example is beta-sialon, where the O and Si are partially replaced by N and Al, respectively. Advanced sialons are now being researched to enhance fracture toughness and improved creep properties.

Aluminides. These are intermetallic compounds of aluminum. The potential of these products include uses where low weight, high-temperature strength, and oxidation resistance are required. Traditionally, these products are made by way of powder metallurgy technology. Powder consolidation has been affected by sintering and hot isostatic pressing, both methods requiring long processing at height temperature. They rely mainly on solid-state diffusion. In a more recent method, *dynamic consolidation* uses high-pressure shock waves traveling at several kilometers per second. Such shocks can be generated through the use of detonating explosives or a gun-fired projectile. Upon full development of the shock-wave technique, advantages predicted include: (1) the nonequilibrium microstructures produced in rapid-solidification processing of powders will be retained in the final compact, (2) composite materials may be fabricated with very thin reaction zones between matrix and reinforcement, thus minimizing brittle reaction products that distract from the composite properties, and (3) net shapes may be produced. Normally confined in the past to production of centimer-size parts, an improved process may be scaled up to meter-size products. Further development is required to prevent the formation of cracks.

Shape-Memory Alloys. Stoeckel defines a shape-memory alloy as the ability of some plastically deformed metals (and plastics) to resume their original shape upon heating. This effect has been observed in numerous metal alloys, notably the Ni-Ti and copper-based alloys, where commercial utilization of this effect has been exploited. (An example is valve springs that respond automatically to change in transmission-fluid temperature.) Copper-based alloy systems also exhibit this effect. These have been Cu-Zn-Al and Cu-Al-Ni systems. In fact, the first thermal actuator to utilize this effect (a greenhouse window opener) uses a Cu-Zn-Al spring.

ARALL Laminates. Developed in the late 1970s, *AR*amid *AL*uminum *L*aminates were developed by Delft University and Fokker Aircraft Co. The laminate currently is used for the skin of the cargo door for the Douglas C-17 military transport aircraft, but additional aerospace applications are envisioned. In essence, the laminate comprises a prepreg (i.e., unidirectional aramid fibers embedded in a structural epoxy adhesive) sandwiched between layers of aircraft alluminum alloy sheet. The fibers are oriented parallel to the rolling direction of the aluminum sheet. Prior to lay-up and autoclave curing, the aluminum surfaces are anodized and primed to ensure good bond integrity and to inhibit corrosion of the metal in the event of moisture intrusion at the bond line.

Quasicrystals. In the early 1980s, D. Schechtman at NIST (U.S. National Institute for Standards and Technology) discovered quasicrystals in aluminum alloys. Since then, they also have been noted in other alloys, including those of copper, magnesium, and zinc. Quasicrystals contradict the traditional fundamentals of crystallography to the effect that the periodicity of a perfect crystal structure is not possible with pentagon shapes. Much pioneering research on quasicrystals also has been conducted at the Laboratoire de Science at Génie des Matériaux Métalliques in France.

To date, little use has been found for quasicrystals in bulk, but they have proved very effective as coatings, notably in cookware. Recent cookware, with a different appearance and "feel," has appeared in the marketplace. These pots, pans, and so on, have a hardness equal to that of hardened alloy steel and thus are practically immune to scratching. They also are thermally stable and corrosion and oxidation resistant.

The coating is applied by using flame, supersonic, and plasma-arc spraying. The deposited material consists of a mixture of quasicrystals and crystalline phases. The quasicrystal content of the surface ranges from 30–70%.

In structure, the quasicrystal relates to the Penrose tile structures (polygon), originally proposed by Roger Penrose, a mathematician at Oxford University. See **Crystal.**

Advances in Powdered Metallurgy (PM) Aluminum Alloys. As noted by Frazier, materials for advanced airframe structures and propulsion systems must withstand increasingly high temperature exposure. For example, frictional heating can raise supersonic skin temperatures to a range of 555° to 625°F (290° to 330°C). Unfortunately, wrought age-hardening aluminum alloys lose strength above 265°F (130°C). Titanium alloys perform well under these conditions, but they are 67% denser than aluminum, constituting about 42% of the weight of contemporary turbofan engines. Replacement of half the titanium with aluminum would reduce engine weight by about 20%. The motivation for using PM products is cost reduction and improved performance. Advanced thermoplastic matrix composites under development are difficult to process and presently cost prohibitive. Thus, intensive research is underway to improve rapid solidification technology and other new PM processes to increase the alloy aluminum content, thus reducing weight and cost.

Aluminum Electroplating. Electroplated aluminum is growing in acceptance for use in automotive parts, electrical equipment, and appliances and for products in a marine environment. Markets may be extended as the result of a new galvano-aluminum electroplating process developed by Siemens Research Laboratory (Erlangen, Germany) and described in the Hans reference.

Additional Reading

Aluminum Association: "Aluminum Standards and Data" and "Aluminum Statistical Review" (issued periodically).

Carter, G. F., and D. E. Paul: "Materials Science and Engineering," ASM International, Materials Park, Ohio, 1991.

Cathonet, P.: "Quasicrystals at Home on the Range," *Adv. Mat. & Proc.*, 6 (June 1991).

Frazier, W. E.: "PM Al Alloys: Hot Prospects for Aerospace Applications," *Adv. Mat. & Proc.*, 42 (November 1988).

Frick, J., Editor: "Woldman's Engineering Alloys," 7th Edition, ASM International, Materials Park, Ohio, 1990.

Gregory, M. A.: "ARALL Laminates Take Wing," *Adv. Mat. & Proc.*, 115 (April 1990).

Hans, R.: "High-Purity Aluminum Electroplating," *Adv. Mat. & Proc*, 14 (June 1989).

Kennedy, D. O.: "SiC Particles Beef up Investment-Cast Aluminum," *Adv. Mat. & Proc.*, 42–46 (June 1991).

Perry, R. H., and D. Green: "Perry's Chemical Engineers' Handbook," 6th Edition, McGraw-Hill, New York, 1990.

Peterson, W. S., Editor.: "Hall-Heroult Centennial—First Century of Aluminum Process Technology—1886–1986," The Metallurgical Society, London, 1986.

Rioja, R. J., and R. H. Graham: "Al-Li Alloys Find Their Niche," *Adv. Mat. & Pro.*, 23 (June 1992).

Samuels, L. E.: "Metals Engineering: A Technical Guide," ASM International, Materials Park, Ohio, 1988.

Sousa, L. J.: "The Changing World of Metals," *Adv. Mat. & Proc.*, 27 (September 1988).

Staff: "Aluminum, Steel Cans Make a Dent in the Market," *Adv. Mat. & Proc.*, 12 (June 1989).

Staff: "Sialons Produced by Combustion Synthesis," *Adv. Mat. & Proc.*. 11 (September 1989).

Staff: "Aluminum Data Sheets, #7450G," ASM International, Materials Park, Ohio, 1990.

Staff: "Strength (Metals)," *Adv. Mat. & Proc.*, 19 (June 1990).

Staff: "Properties and Selection: Nonferrous Alloys and Special-Purpose Materials," ASM International, Materials Park, Ohio, 1991.

Staff: "Audi To Get Aluminum Space Frame," *Adv. Mat. & Proc.*, 9 (January 1992).

Staff: "Forecast '92—Aluminum," *Adv. Mat. & Proc.*, 17 (January 1992).

Staff: "Handbook of Chemistry and Physics," 73rd Edition, CRC Press, Boca Raton, Florida, 1992–1993.

Stoeckel, D.: "Shape-Memory Alloys Prompt New Actuator Designs," *Adv. Mat. & Proc.*, 33 (October 1990).

Strauss, S.: "Impossible Matter (Quasicrystals)," *Techy. Review (MIT)*, 19 (January 1991).

Taketani, H.: "Properties of Al-Li Alloy 2091-T3 Sheet," *Adv. Mat. & Proc.*, 113 (April 1990).

Van Horn, K. R., Editor: "Aluminum," Vol. 1–3, ASM International, Materials Park, Ohio, 1967. (A classic reference.)

Vassilou, M. S.: "Shock Waves Shape Aluminides," *Adv. Mat. & Proc.*, 70 (October 1990).

Vaughan, D. E. W.: "The Synthesis and Manufacture of Zeolites," *Chem. Eng. Prog.*, 25 (February 1988).

Webster, D., Haynes, T. G., III, and R. H. Fleming: "Al-Li Investment Castings Coming of Age," *Adv. Mat. & Proc.*, 25 (June 1988).

Webster, D., and C. G. Bennett: "Tough(er) Aluminum-Lithium Alloys," *Adv. Mat. & Proc.*, 49 (October 1989).

Winterbottom, W. L.: "The Aluminum Auto Radiator Comes of Age," *Adv. Mat. & Proc.*, 55 (May 1990).

ALUNITE. The mineral alunite, $KAl_3(SO_4)_2(OH)_6$, is a basic hydrous sulfate of aluminum and potassium; a variety called natroalunite is rich in soda. Alunite crystallizes in the hexagonal system and forms rhombohedrons with small angles, hence resembling cubes. It may be in fibrous or tabular forms, or massive. Hardness, 3.5–4; sp gr, 2.58–2.75; luster, vitreous to pearly; streak white; transparent to opaque; brittle; color, white to grayish or reddish.

Alunite is commonly associated with acid lavas due to sulfuric vapors often present; it may occur around fumaroles or be associated with sulfide ore bodies. It has been used as a source of potash. Alunite is found in the Czech Republic and Slovakia, Italy, France, and Mexico; in the United States, in Colorado, Nevada, and Utah.

Alunite is also known as *alumstone*.

ALVEOLUS. 1. A minute sac-like chamber in a hollow organ, such as the air sacs of the lungs and the components of various glands. Its sac-like form distinguishes it from other chambers in which the walls are relatively thicker. 2. The cavity or socket in the jaw in which the root of a tooth is fixed.

ALZHEIMER'S DISEASE AND OTHER DEMENTIAS. Today, dementia is most commonly associated with Alzheimer's disease (AD), although there are other neurologic disorders that cause dementia. The latter are described briefly towards the end of this article.

Only within the relatively recent past has AD gained formal medical recognition and prominence. It does seem reasonable, however, to assume that AD has been a major cause of dementia, particularly among the elderly, for a long time, possibly dating back to the dawn of humankind. Little research has been directed toward fixing the origin or present demographics of the disease, with the exception of a report prepared by a Dr. Alois Alzheimer (Germany) in 1907 concerning the symptoms of the disease. This report will be described a bit later in this article.

Some reporters have called AD an epidemic, with its estimated 4 million cases in the United States alone and with several hundred thousand new cases reported annually. These large figures are more of a statistical aberration than they are the suggestion of an epidemic, simply because the reporting of cases commenced just a few years ago. Nevertheless, the AD case load is staggering.

The period from first detection of AD to termination ranges from a few years up to 15 years or more, during which time the patient requires progressively increasing care. Ultimately, there is total loss of mental function and control of bodily functions, representing great family anguish and high medical costs.

As with other diseases, AD can be defined in terms of (1) symptoms, and (2) the biological processes that bring about the disease. Considering the present state of AD research, the actual cause (etiology) of the disease cannot be stated definitively. Present AD research is directed toward determining the cause. However, the end condition (at death of patient) can be determined by an autopsy of the brain. To date, only limited success has been achieved by using noninvasive instrumental means, such as the CAT scan, to learn of the progress of the disease in the living patient.

AD Symptoms. It is interesting to note that, in retrospect, a report prepared by Alois Alzheimer in 1907 rather accurately described the symptoms of an AD patient in his care. His description of the unnamed disease at that time was of little significance for nearly a century. In his report, typical AD symptoms were described and included:

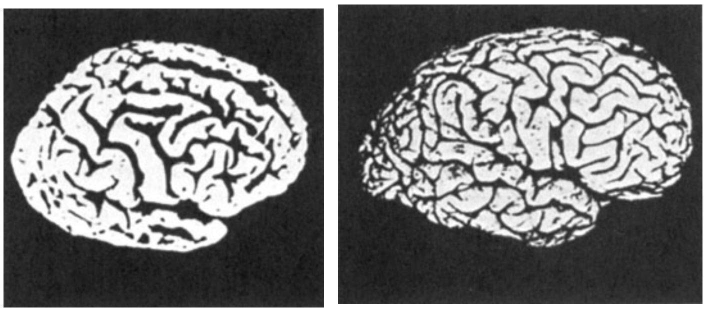
1. Unexplainable alteration of personality, including fits of jealousy, fear of the unknown, and fear of being harmed or murdered;
2. Outbursts of imagined terror;
3. Random but progressively occurring periods of forgetfulness;
4. Seriously erratic behavior, such as moving household objects about and sometimes hiding them for no particular reason; and
5. Getting "lost" in one's own household, not knowing the time of day, and like factors.

Some of these patterns, of course, are present in other mental illnesses, and all are not apparent in any given AD patient.

To assist in differentially diagnosing AD, a work group made up of scientists with the *National Institute of Neurological and Communicative Disorders and Strokes* and the *Alzheimer's Disease and Related Disorders* organizations has suggested a variety of criteria for pinpointing an AD patient from others who may exhibit some AD symptoms, but who instead are suffering the effects of strokes, depression, or drug reactions. In comparing AD patient behavior with later findings by autopsy, a success rate of up to 92% has been claimed.

As with other diseases that usually are terminal, the question of telling or not telling the AD patient of the condition is debatable. In an excellent article by Drickamer and Lachs (see Additional Reading), the pros and cons are discussed and can be helpful to the physician and family members who must face this problem.

AD Biological Process(s). What does the examiner find when an autopsy of the brain is conducted?



Brain comparison: (*Left*) Brain from an Alzheimer's disease patient; (*Right*) Normal brain. (*Laboratory of Neuroscience, National Institute of Aging.*)

1. The brain of an AD patient may be smaller and of a coarser appearance. (See accompanying figure.)
2. Plaques of the brain protein (beta amyloid) that are surrounded by neurons, the latter featuring neurofibrillary tangles.
3. A loss of neuronal cells.
4. Vascular damage resulting from the deposition of plaque.

A more technically worded definition is contained in the Murrell, et al. (1981) article (*see Additional Reading*): "Alzheimer's disease is a form of localized amyloidosis characterized by cerebral cortical amyloid plaques, neurofibrillary tangles, and amyloid deposits within the walls of leptomeningeal vessels."

Avenues of Research. A clear-cut tie of a specific gene mutant that can be passed along to family progeny would, as with a few genetic diseases researched thus far, would have provided (perhaps) a relatively fast answer to AD research. It has been found that with some AD patients a familial manifestation of the disease occurs, and research along these lines is being conducted, as indicated by the Murrell, et al. (1991) article previously mentioned. But hereditary or familial AD represents only some of the cases. Research to date indicates that a majority of AD cases are *sporadic* (nonfamilial).

The key research target of many AD research groups has centered on the protein beta-amyloid, its precursor (APP), and related chemistry, including degenerative products. Nerve growth factor also has been a target.

Because of the inability to observe the progress of AD in living patients, a number of researchers have developed transgenic animals (mouse and rat) that will exhibit Alzheimer's-like pathology. After claiming success in this area in 1991, some researchers have more recently expressed disappointment with the animal models developed.

Researchers to date have not found the fundamental key(s) to AD, and the prospects of solving this complex riddle do not appear to lie within a short time frame. Forecasts for finding the fundamental causation (or multiple factors) range from mildly pessimistic to mildly optimistic. Meanwhile, AD research is enriching the scientific knowledge on the higher functions of the brain, which is the target of Alzheimer's disease.

Arteriosclerotic Dementia

In *arteriosclerotic dementia*, many of the symptoms of which parallel those of Alzheimer's disease, the cause of the condition is well known, namely, a reduced cerebral blood flow. In Alzheimer's disease, the blood flow is normal. Arteriosclerotic dementia is sometimes called *multi-infarct dementia* because of atherosclerosis of small arteries and patchy cerebral infarction. The onset of this condition may be abrupt.

Occult Hydrocephalus

In the syndrome of *occult hydrocephalus*, where there is normal cerebrospinal fluid pressure, some researchers have found that this form of dementia may be reversible as the result of shunting. At one time, it was believed that shunting may improve Alzheimer's disease, but this has not proven so. However, it is difficult for neurosurgeons to discern situations where shunting may or may not be useful. Some neurologists prefer to use the shunt procedure only in cases where there has been a known episode of meningitis, trauma, or subarachnoid hemorrhage.

See also **Brain (Injury);** and **Nervous System and The Brain.**

Additional Reading

Cooke, R.: "Attacking Alzheimer's," *Technology Review (MIT)*, 19 (July 1988).

Cummings, J. L., and B. L. Miller, Editors: *Alzheimer's Disease: Treatment and Longterm Management*, Marcel Dekker, New York, 1990.

Drickamer, M. A., and M. S. Lachs: "Should Patients with Alzheimer's Disease Be Told Their Diagnosis?" *N. Eng. J. Med.*, 947 (April 2, 1992).

Erickson, D.: "Love and Terror: Is a Chemical Messenger Key to Treating Alzheimer's?" *Sci. Amer.*, 148 (April 1991).

Erickson, D.: "Model Mice: Transgenic Animals and Alzheimer's Research," *Sci. Amer.*, (September 1991).

Hardy, J. A., and G. A. Higgins: "Alzheimer's Disease: The Amyloid Cascade Hypothesis," *Science*, 184 (April 10, 1992).

Holloway, M.: "The Estrogen Factor," *Sci. Amer.*, 26 (June 1992).

Kosik, K. S.: "Alzheimer's Disease: A Cell Biological Perspective," *Science*, 780 (May 8, 1992).

Marx, J.: "Brain Protein Yields Clues to Alzheimer's Disease," *Science*, 1664 (March 31, 1989).

Marx, J.: "Mutation Identified as a Possible Cause of Alzheimer's Disease," *Science*, 876 (February 22, 1991).

Marx, J.: "Alzheimer's Research Moves to Mice," *Science*, 266 (July 19, 1991).

Marx, J.: "New Clue Found to Alzheimer's," *Science*, 857 (August 23, 1991).

Marx, J.: "Boring in on Beta-Amyloid's Role in Alzheimer's," *Science*, 688 (February 7, 1992).

Marx, J.: "Potential New Alzheimer's Test," *Science*, 1279 (May 29, 1992).

Murrell, J., et al.: "A Mutation in the Amyloid Precursor Protein Associated with Hereditary Alzheimer's Disease," *Science*, 97 (October 4, 1991).

Relman, A. S.: "Tacrine as a Treatment for Alzheimer's Dementia," *N. Eng. J. Med.*, 349 (January 31, 1991).

Rennie, J.: "The Mice That Missed," *Sci. Amer.*, 20 (June 1992).

Schubert, D., et al.: "Amyloid Beta Protein Precursor is Possibly a Heparin Sulfate Proteoglycan Core Protein," *Science*, 223 (July 8, 1988).

Selkoe, D. J.: "Amyloid Protein and Alzheimer's Disease," *Sci. Amer.*, 68 (November 1991).

Weiss, R.: "Toward a Future with Memory: Researchers Look High and Low for the Essence of Alzheimer's," *Science News*, 120 (February 24, 1990).

AMALGAM. 1. An alloy containing mercury. Amalgams are formed by dissolving other metals in mercury, when combination takes place often with considerable evolution of heat. Amalgams are regarded as compounds of mercury with other metals, or as solutions of such compounds in mercury. It has been demonstrated that products which contain mercury and another metal in atomic proportions may be separated from amalgams. The most commonly encountered amalgams are those of gold and silver. See also **Gold; Mercury; Silver.**

2. A naturally occurring alloy of silver with mercury, also referred to as mercurian silver, silver amalgam, and argental mercury. The natural amalgam crystallizes in the isometric system; hardness, 3–3.5; sp gr, 13.75–14.1; luster, metallic; color, silver-white; streak, silver-white; opaque. Amalgam is found in Bavaria, British Columbia, Chile, the Czech Republic and Slovakia, France, Norway, and Spain. In some areas, it is found in the oxidation zone of silver deposits and as scattered grains in cinnabar ores.

AMARANTHUS. Genus of plants in the family *Amaranthaceae* (amaranth family), including many coarse and rather obnoxious pigweeds. *Amaranthus caudatus*, the love-lies-bleeding; and the *A. hypochondriacus*, the Princess feather are cultivated in gardens for their attractive green, purple, and crimson flowers. They are annuals.

AMBER. Amber is a fossil resin which has been known since early times because of its property of acquiring an electric charge when rubbed. In modern times it has been used largely in the making of beads, cigarette holders, and trinkets. Its amorphous non-brittle nature permits it to be carved easily and to acquire a very smooth and attractive surface. Amber is soluble in various organic solvents, such as ethyl alcohol and ethyl ether.

It occurs in irregular masses showing a conchoidal fracture. Hardness, 2.25; sp gr, 1.09; luster, resinous; color, yellow to reddish or brownish; it may be cloudy. Some varieties will exhibit fluorescence. Amber is transparent to translucent, melts between 250 and 300°C.

Amber has been obtained for over 2,000 years from the lignite-bearing Tertiary sandstones on the coast of the Baltic Sea from Gdańsk to Liepája; also from Denmark, Sweden and the other Baltic countries. Sicily furnishes a brownish-red amber that is fluorescent.

The association of amber with lignite or other fossil woods, as well as the beautifully preserved insects that are occasionally in it, is ample proof of its organic origin.

AMBERGRIS. A fragrant waxy substance formed in the intestine of the sperm whale and sometimes found floating in the sea. It has been used in the manufacture of perfumes to increase the persistence of the scent. See also **Whales, Dolphins,** and **Porpoises.**

AMBIENT CONDITIONS. These are the conditions of environment in which an instrument or device operates. As various environmental factors change, they may seriously alter the accuracy and other performance parameters of equipment. The effects of ambient changes may be combated in two basic ways: (1) *ambient compensation* in which a technique is used to offset the effects of environmental changes, such as changes in temperature; pressure; supply voltage, frequency, and waveform; shock and vibration; and position, and (2) *ambient protection* in which the equipment is protected from its environment, that is, by way of isolation, insulation, and similar techniques. When procuring instruments and other sensitive laboratory and research equipment, the expected range of ambient conditions over which the equipment will be used always should be specified, and the manufacturer's assurance of ambient compensation and/or protection should be studied carefully.

Ambient Compensation. An electrical circuit may be made relatively insensitive to temperature changes by use of components, such as *manganin* or *evanohm* wire for resistors. The resistances of these alloys are only slightly affected by temperature changes in the range of laboratory and industrial environments. Mechanical means of offsetting temperature changes include bimetals or other combinations of materials with different temperature coefficients of expansion. Where stability of dimensions must be maintained, materials with very low thermal expansion coefficients, such as *invar*, may be used in construction of the equipment.

The spring rate of a pressure-sensitive element, such as a bourdon tube, will be affected by temperature change. For most materials, the modulus of elasticity will decrease with increasing temperature, thereby causing an incorrect higher indication. Use of a bimetal to reduce the span is one common means for correcting this effect. Making the pressure element of *Ni-span C*, which has an essentially constant modulus, is another way of correcting the condition.

Manual temperature correction means are provided on several instruments, such as mercurial barometers and precision manometers. Volume changes in a fluid-filled device caused by temperature variations may be offset by a pressure-sensitive diaphragm, which is deflected by expansion of the fluid to maintain a relatively constant pressure. A similar principle is used in the hydraulic accumulator.

To prevent a pressure-sensitive element from being affected by ambient pressure changes, the element may be hermetically sealed within a rigid enclosure.

Correction curves and tables also are widely used. The output of the uncompensated instrument is read, the ambient conditions noted, and the proper correction found in a table. This technique is less convenient than automatic or even manual compensation. However, under reasonably stable environmental influences, the method may be more accurate and practical than automatic compensation.

Instruments are usually designed to tolerate some variation in supply voltage. The availability of zener diodes in a wide range of voltage ratings simplifies the problem of maintaining dc voltage levels with almost any required degree of exactness. Power supplies with ac line input and dc output are available commercially with almost any desired combination of voltage, current, range of adjustability, line and load regulation, ambient temperature compensation, ripple, transient suppression, and package configuration. Alternating current voltage regulators can be obtained to maintain constant output within the needed tolerance for input variations of any likely magnitude, with widely differing speeds of response and power capabilities. The quality of the output waveforms differs markedly, depending upon the operating principle and the magnitude and nature of the load. Care always must be exercised in determining the quality of waveform needed to assure satisfactory operation of the equipment being powered. See also **Power Sources and Supplies.**

The accuracy and stability of line power frequency normally are adequate for most applications. Short-term frequency variations seldom are more than a fraction of a hertz on a 50- or 60-Hz line. Where exceptionally close regulation of frequency is needed, as in the case of very accurate timing, an electronic oscillator may be built into the equipment.

Ambient Protection. Ambient condition protection may be used to isolate the instrument or device from the undesirable environment. The dividing line between this approach and that of ambient compensation is narrow and sometimes difficult to define.

Cleanliness of environment, such as air or liquid as an input to equipment, is an ambient condition difficult to measure, but often necessary for satisfactory long-term operation. Where gases and liquids are involved, it may be necessary to remove solid particles by mechanical filtering, washing, or centrifuging; moisture may have to be removed by cooling and condensing, by the use of desiccants, or by centrifuging; oil may have to be removed by chemical absorption or mechanical filtering. These problems pertain to both signal and control lines in connection with pneumatic equipment. Further, the instrument box may have to be protected from dust, oil, and chemicals in the atmosphere. Where it is not practical to seal the instrument enclosure sufficiently tightly to exclude foreign materials, purging may be an attractive alternative approach. Here, the air inside the instrument case is maintained at a pressure slightly above that of the outside environment. The technique also can be used to protect electrical equipment from hazards caused by combustible gases or vapors entering the enclosure where heat or a spark might cause an explosion. See also **Intrinsic Safety.**

Changes in environmental humidity affect electrical leakage, dimensions of hygroscopic materials such as paper-recorded charts, and the concentration of chemical solutions, and, of course, the presence of dampness enhances the formation of rust and corrosion. Methods for overcoming humidity effects include sealing to exclude moisture, heating the equipment (maintaining a constant temperature) to lower the relative humidity, cooling by refrigeration to cause condensation on noncritical areas and permitting condensate removal, and the use of desiccants to absorb moisture. Pipes and tubing containing fluids that would freeze or become too viscous when subjected to very low ambient temperatures can be "steam traced" by paralleling the tubing with a small steam line.

Sometimes a practical solution for assuring satisfactory operation despite extremes of temperature is that of derating a component, so that its output will be adequate under the most unfavorable conditions. This applies particularly to electronic components, such as transistors, whose power dissipation decreases as the temperature rises.

The best protection against shock and vibration is careful design of the original equipment. Serious effects on the performance of transformers, capacitors, and resistors may occur unless relative motion of the materials used in constructing them is prevented, as well as motion of the entire component relative to the remainder of the circuit. Heavy parts, such as transformers, must be rigidly attached to a sufficiently rugged chassis to prevent them from breaking loose. Where shock and vibration conditions are contemplated, expensive instruments should be subjected to prescribed vibration and shock testing prior to acceptance.

Grounding may be important both for low level signals (to minimize electrical noise problems) and for high power circuits (to prevent hazards to personnel). Particular note should be taken of the distinction between line power neutral and true ground. Even a small resistance between them can produce an unacceptably high voltage difference, if there is a significant current. In many low-level instrument systems, it is vital to proper operation that a common ground or shield be used for all components of the system. This can be even more important than assuring that the ground is at true earth potential. See also **Common-Mode Voltage; Differential-Mode Voltage.**

Frequently, an expendable means may be used for protection, as exemplified by a fuse in an electrical circuit. Refinements include fast- and slow-blow fuses, and circuit breakers with various speeds of operation. Voltage limiting means are the neon lamp and the zener diode, which will pass essentially no current unless the voltage (to ground, or some other safe part of the circuit) exceeds the allowable value.

Hydraulic and pneumatic equivalents are rupture disks, which break to prevent pressure from exceeding a predetermined limit, and "hydraulic fuses," which are preloaded valves that close if the flow rate becomes too great.

Thermal devices with similar purposes are fusible links, as found in automatic sprinkler systems, and freeze-out plugs, which release if water freezes and expands, endangering a system.

AMBLYGONITE. A rather rare compound of fluorine, lithium, aluminum, and phosphorus, (Li, Na)$AlPO_4$(F, OH). It crystallizes in the tri-clinic system; hardness, 5–5.6; sp gr 3.08; luster, vitreous to greasy or pearly; color, white to greenish, bluish, a yellowish or grayish; streak white; translucent to subtransparent.

Amblygonite occurs in pegmatite dikes and veins associated with other lithium minerals. It is used as a source of lithium salts. The name is derived from two Greek words meaning blunt and angle, in reference to its cleavage angle of 75° 30′.

Amblygonite is found in Saxony; France; Australia; Brazil; Varutrask, Sweden; Karibibe, S.W. Africa; and in the United States at Hebron, Paris, Greenwood, Rumford and Auburn, Maine; Branchville, Connecticut; Black Hills, South Dakota; and Pala, California.

AMBLYOPIA. See **Vision and the Eye.**

AMBUSH BUG (*Insecta, Hemiptera*). Predacious bugs named for their habit of lying in wait for their prey in flowers, where their colors conceal them. They belong to the family *Phymatidae*. The forelegs of these predatory bugs are prehensile; the tibia can be folded back against the femur, and the tarsus is regressed. Most species are tropical, only two having penetrated as far as Europe and only one of these (*Phymata crassipes*) as far north as Germany. In the United States, the ambush bug (*Phymata fasciata*), a yellowish ambush bug which may be up to 12 millimeters (0.5 inch) long, is the most commonly encountered species, though other species are also found (for example, *Phymata erosa*).

AMEBIASIS (Amebic Dysentery). An invasion of the intestinal tract by the protozoan *Entameba histolytica* which exists in two forms: (1) the motile, phagocytic trophozoite (10-20 micrometers in diameter), which posses a single nucleus and granular cytoplasm. Passed in loose stools, the trophozoite is not hardy and does not survive outside the body. The cyst phase which is derived from the trophozoite in the colon and has one to four nuclei can, however, survive environmental stress as well as passage through the stomach and is primarily responsible for human infestation. The infection is derived from fecal-oral, water, or foodborne routes as well as by person-to-person transmission. The latter accounts for the heightened prevalence of *E. histolytica* infection in homosexuals and in institutions where there is fecal incontinence and poor hygiene. About 10% of the world's population is infected with the protozoa.

The trophozoites, being cytolytic, may invade the bowel wall and produce ulcerations; locales most frequently affected are the cecum and sigmoid colon. In asymptomatic chronic infection, a granuloma (ameboma) can develop at a site of infection.

Where the infestation is not quiescent, the illness can produce episodes of diarrhea or even fulminant dysentery. Abdominal distress, tenesmus, dull sacral pain and flatulence are common. Amebiasis may spread hematogenously to any body organ, the liver being most commonly infected and then the lungs. In the liver, necrosis results in the formation of abscesses ranging in size from a few centimeters to 20 centimeters in diameter. Ninety percent of the abscesses are seen in the right lobe, with males being seven times more likely to be affected than females. Rupture of right lobe abscesses can result in extension of the infection to the chest or peritoneal cavity; left lobe rupture can extend to the pericardium with fatal results.

Presenting symptoms in amebic liver abscess include malaise, fatigue, anorexia, dull aching abdominal pain, localized in the right upper quadrant or right chest.

Cutaneous amebiasis, most commonly seen on the perineum or genitalia can result in ulcerative or fungating lesions arising from trophozoite invasion through fecal or sexual transmission.

Apart from the physical signs and symptoms, diagnosis of amebiasis stems from microscopic examination of feces, care being taken to distinguish between *E. histolytica* bodies and those of *Entameba coli* and *Campylobacter*.

Metronidazole (*Flagyl*®) is usually the most effective treatment but may be substituted by Diloxamide furoate when side effects predominate.

A. C. Vickery, Ph.D.; Associate Professor, College of Public Health, University of South Florida, Tampa, Florida.

AMEIVA (*Reptilia, Sauria*). A lizard of Central and South America, small, very active, timid, and mainly insectivorous. The name is that of the genus, applied as a common name to the score of included species.

AMENORRHEA. The absence of menstruation, which may be primary or secondary. Secondary amenorrhea is defined as the absence of menstruation in a woman who has had previously normal menstrual periods. The most obvious cause of this, of course, is pregnancy. The physician will probe into four possible areas of dysfunction in attempting to comprehend a particular primary or secondary amenorrhea—hypothalamic, pituitary, ovarian, and uterine defects. See **Gonads;** and **Hormones.**

AMERICIUM. Chemical element, symbol Am, at no. 95, at. wt. 243 (mass number of the most stable isotope), radioactive metal of the actinide series, also one of the transuranium elements. All isotopes of americium are radioactive; all must be produced synthetically. The element was discovered by G. T. Seaborg and associates at the Metallurgical Laboratory of the University of Chicago in 1945. At that time, the element was obtained by bombarding uranium-238 with helium ions to produce ^{241}Am which has a half-life of 475 years. Subsequently, ^{241}Am has been produced by bombardment of plutonium-241 with neutrons in a nuclear reactor. ^{243}Am is the most stable isotope, an alpha emitter with a half-life of 7950 years. Other known isotopes are ^{237}Am, ^{238}Am, ^{240}Am, ^{241}Am, ^{242}Am, ^{244}Am, ^{245}Am, and ^{246}Am. Electronic configuration is $1s^22s^22p^63s^23p^63d^{10}\ 4s^24p^64d^{10}\ 4f^{14}\ 5s^25p^65d^{10}5f^76s^26p^67s^2$. Ionic radii are: Am^{4+}, 0.85 Å; Am^{3+}, 1.00 Å.

This element exists in acidic aqueous solution in the (III), (IV), (V), and (VI) oxidation states with the ionic species probably corresponding to Am^{3+}, Am^{4+}, AmO_2^+ and AmO_2^{2+}. The oxidation potentials in acidic aqueous solution are summarized in the following diagram in which the americium(IV)-americium(V) and americium(III)-americium(V) couples are calculated from the others.

$$\begin{tikzcd}
Am \arrow[r, "<2.7\,V"] \arrow[rr, bend right, "+2.32\,V"'] & Am^{+2} \arrow[r, ">1.5\,V"] & Am^{+3} \arrow[r, "-2.7\,V"] \arrow[rr, bend left, "(-1.5\,V)"] \arrow[rrr, bend right, "-1.74\,V"'] & Am^{+4} \arrow[r, "(-1.04\,V)"] & AmO_2^{+} \arrow[r, "-1.60\,V"] & AmO_2^{+2}
\end{tikzcd}$$

The colors of the ions are: Am^{3+}, pink; Am^{4+}, rose; AmO_2^+, yellow; and AmO_2^{2+}, rum-colored.

It can be seen that the (III) state is highly stable with respect to disproportionation in aqueous solution and is extremely difficult to oxidize or reduce. There is evidence for the existence of the (II) state since tracer amounts of americium have been reduced by sodium amalgam and precipitated with barium chloride or europium sulfate as carrier. The (IV) state is very unstable in solution: the potential for americium(III)-americium(IV) was determined by thermal measurements involving solid AmO_2. Americium can be oxidized to the (V) or (VI) state with strong oxidizing agents, and the potential for the americium(V)-americium(VI) couple was determined potentiometrically.

In its precipitation reactions americium(III) is very similar to the other tripositive actinide elements and to the rare earth elements. Thus the fluoride and the oxalate are insoluble and the phosphate and iodate are only moderately soluble in acid solution, whereas the nitrates, halides, sulfates, sulfides, and perchlorates are all soluble. Americium(VI) can be precipitated with sodium acetate giving crystals isostructural with sodium uranyl acetate,

$$NaUO_2(C_2H_3O_2)_3 \cdot xH_2O$$

and the corresponding neptunium and plutonium compounds.

Of the hydrides of americium, both AmH_2 and Am_4H_{15} are black and cubic.

When americium is precipitated as the insoluble hydroxide from aqueous solution and heated in air, a black oxide is formed which corresponds almost exactly to the formula AmO_2. This may be reduced to Am_2O_3 through the action of hydrogen at elevated temperatures. The AmO_2 has the cubic fluorite type structure, isostructural with UO_2, NpO_2, and PuO_2. The sesquioxide, Am_2O_3 is allotropic, existing in a reddish brown and a tan form, both hexagonal. As in the case of the preceding actinide elements, oxides of variable composition between $AmO_{1.5}$ and AmO_2 are formed depending upon the conditions.

All four of the trihalides of americium have been prepared and identified. These are prepared by methods which are similar to those used in the preparation of the trihalides of other actinide elements. AmF_3 is pink and hexagonal, as is $AmCl_3$; $AmBr_3$ is white and orthorhombic; while a tetrafluoride, AmF_4 is tan and monoclinic.

In research at the Institute of Radiochemistry, Karlsruhe, West Germany during the early 1970s, investigators prepared alloys of americium with platinum, palladium, and iridium. These alloys were prepared by hydrogen reduction of the americium oxide in the presence of finely divided noble metals according to:

$$Am_2O_3 + 10\,Pt \xrightarrow[1100°C]{H_2} 2\,AmPt_5 + H_2O$$

The reaction is called a *coupled reaction* because the reduction of the metal oxide can be done only in the presence of noble metals. The hydrogen must be extremely pure, with an oxygen content of less than 10^{-25} torr.

See also **Chemical Elements.**

Industrial utilization of americium has been quite limited. Uses include a portable source for gamma radiography, a radioactive glass thickness gage for the flat glass industry, and an ionization source for smoke detectors.

Americium is present in significant quantities in spent nuclear reactor fuel and poses a threat to the environment. A group of scientists at the U.S. Geological Survey (Denver, Colorado) has studied the chemical speciation of actinium (and neptunium) in ground waters associated with rock types that have been proposed as possible hosts for nuclear waste repositories. Researchers Cleveland, Nash, and Rees (see reference list) concluded that americium (and neptunium) are relatively insoluble in ground waters containing high sulfate concentrations (90°C). The insoluble neptunium species is Np(IV); hence reducing waters should enhance its formation. Americium can exist only in the trivalent state under experimental conditions. The solubility of americium should be representative of that of curium.

Additional Reading

Asprey, L. B., Stephanou, W. E., and R. A. Penneman: "Hexavalent Americium." *Amer. Chem. Soc. J.*, **73**, 5715–5717 (1951).

Asprey, L. B.: "New Compounds of Quadrivalent Americium, AmF_4, $KAmF_5$," *Amer. Chem. Soc. J.*, **76**, 2019–2020 (1954).

Cleveland, J. M., Nash, K. L., and T. F. Rees: "Neptunium and Americium Speciation in Selected Basalt, Granite, Shale, and Tuff Ground Waters," *Science*, **221**, 271–273 (1983).

Cunningham, B. B.: "Chemistry of Element 96," Metallurgical Laboratory Report CS-3312, Univ. of Chicago, Chicago, 1945.

Fisk, Z., et al.: "Heavy-Electron Metals: New Highly Correlated States of Matter," *Science*, **33** (January 1, 1988).

Ghiorso, A., James, R. A., Morgan, L. O., and G. T. Seaborg: "Preparation of Transplutonium Isotopes by Neutron Irradiation," *Phys. Rev.*, **78**, 4, 472 (1950).

Graf, P., et al.: "Crystal Structure and Magnetic Susceptibility of American Metal," *Amer. Chem. Soc. J.*, **78**, 2340 (1956).

Hammond, C. R.: "The Elements" in "Handbook of Chemistry and Physics," 67th Edition, CRC Press, Boca Raton, Florida (1986–1987).

Keller, C., and B. Erdmann: "Preparation and Properties of Transuranium Element—Noble Metal Alloy Phases," *Proc. of the Moscow Symposium on Chemistry of Transuranium Elements*, Pergamon, Elmsford, New York, 1976.

Marks, T. J.: "Actinide Organometallic Chemistry," *Science*, **217**, 989–997 (1982).

Roof, R. B., et al.: "High-Pressure Phase in Americium Metal," *Science*, **207**, 1353–1354 (1980).
Seaborg, G. T.: "The Chemical and Radioactive Properties of the Heavy Elements." *Chemical & Engineering News*, **23**, 2190–2193 (1945).
Seaborg, G. T., Editor: "Transuranium Elements," Dowden, Hutchinson & Ross, Stroudsburg, Pennsylvania, 1978.
Smith, J. L., and R. G. Haire: "Superconductivity of Americium," *Science*, **200**, 535–537 (1978).
Staff: "Handbook of Chemistry and Physics," 73rd Ed., CRC Press, Boca Raton, Florida (1992–1993).

AMETHYST. A purple- or violet-colored quartz having the same physical characteristics as quartz. The source of color is not definite but thought to be caused by ferric iron contamination. Oriental amethysts are purple corundum.

Amethysts are found in the Ural Mountains, India, Sri Lanka, Madagascar, Uruguay, Brazil, the Thunder Bay district of Lake Superior in Ontario, and Nova Scotia; in the United States, in Michigan, Virginia, North Carolina, Montana, and Maine.

The name amethyst is generally supposed to have been derived from the Greek word meaning not drunken. Pliny suggested that the term was applied because the amethyst approaches but is not quite the equivalent of a wine color.

See also **Quartz.**

AMICRON. A name applied by Zsigmondy to individual disperse particles invisible under the ultramicroscope whose size is about 10^{-7} cm. They act as nuclei for the formation of submicrons which are about five times as large.

AMIDES. An amide may be defined as a compound that contains the $CO{\cdot}NH_2$ radical, or an acid radical(s) substituted for one or more of the hydrogen atoms of an ammonia molecule. Amides may be classified as (1) *primary amides* which contain one acyl radical, such as $—CO{\cdot}CH_3$ (acetyl) or $—CO{\cdot}C_6H_5$ (benzoyl), linked to the *amido* group ($—NH_2$). Thus, acetamide NH_2COCH_3 is a combination of the acetyl and amido groups; (2) *secondary amides* which contain two acyl radicals and the *imido* group ($=NH_2$). Diacetamide $HN(COCH_3)_2$ is an example; and (3) *tertiary amides* which contain three acyl radicals attached to the N atom. Triacetamide $N(COCH_3)_3$ is an example.

A further structural analysis will show that amides may be regarded as derivatives of corresponding acids in which the amido group substitutes for the hydroxyl radical OH of the carboxylic group COOH. Thus, in the instance of formic acid HCOOH, the amide is $HCOONH_2$ (formamide); or in the case of acetic acid CH_3COOH, the amide is CH_3CONH_2 (acetamide). Similarly, urea may be regarded as the amide of carbonic acid (theoretical) O:C, that is, NH_2CONH_2 (urea). The latter represents a dibasic acid in which two H atoms of the hydroxyl groups have been replaced by amido groups. A similar instance, malamide,

$$NH_2CO{\cdot}CH_2CH(OH){\cdot}CONH_2$$

is derived from the dibasic acid, malic acid,

$$OHCO{\cdot}CH_2CH(OH){\cdot}COOH.$$

Aromatic amides, sometimes referred to as *arylamides*, exhibit the same relationship. Note the relationship of benzoic acid C_6H_5COOH with benzamide $C_6H_5CONH_2$. *Thiamides* are derived from amides in which there is substitution of the O atom by a sulfur atom. Thus, acetamide $NH_2{\cdot}CO{\cdot}CH_3$, becomes thiacetamide $NH_2{\cdot}CS{\cdot}CH_3$; or acetanilide $C_6H_5{\cdot}NH{\cdot}CO{\cdot}CH_3$ becomes thiacetanilide $C_6H_5{\cdot}NH{\cdot}CS{\cdot}CH_3$. *Sulfonamides* are derived from the sulfonic acids. Thus, benzene-sulfonic acid $C_6H_5{\cdot}SO_2{\cdot}OH$ becomes benzene-sulfonamide $C_6H_5{\cdot}SO_2{\cdot}NH_2$. See also **Sulfonamide Drugs.**

Amides may be made in a number of ways. Prominent among them is the acylation of amines. The agents commonly used are, in order of reactivity, the acid halides, acid anhydrides, and esters. Such reactions are:

$$R'COCl + HNR_2 \rightarrow R'C(=O)NR_2 + HCl$$
$$R'C(=O)OC(=O)R' + HNR_2 \rightarrow R'C(=O)NR_2 + R'COOH$$
$$R'C(=O)OR'' + HNR_2 \rightarrow R'C(=O)NR_2 + R''OH$$

The hydrolysis of nitriles also yields amides:

$$RCN + H_2O \xrightarrow{OH} RCONH_2$$

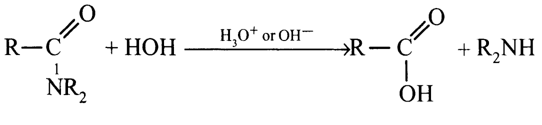
Amides are resonance compounds, having an ionic structure for one form:

$$R—C(=O)NR_2 \qquad R—C(—O^-){:}N^+R_2$$

Evidence for the ionic form is provided by the fact that the carbon-nitrogen bond (1.38 Å) is shorter than a normal C—N bond (1.47 Å) and the carbon-oxygen bond (1.28 Å) is longer than a typical carbonyl bond (1.21 Å). That is, the carbon-nitrogen bond is neither a real C—N single bond nor a C=N double bond.

The amides are sharp-melting crystalline compounds and make good derivatives for any of the acyl classes of compounds, i.e., esters, acids, acid halides, anhydrides, and lactones.

Amides undergo hydrolysis upon refluxing in H_2O. The reaction is catalyzed by acid or alkali.



$$R—C(=O)NR_2 + HOH \xrightarrow{H_3O^+ \text{ or } OH^-} R—C(=O)OH + R_2NH$$

Primary amides may be dehydrated to yield nitriles.

$$R—CONH_2 + C_6H_5SO_2Cl \xrightarrow[70^\circ]{\text{pyridine}} R—CN + C_6H_5SO_3H + HCl$$

The reaction is run in pyridine solutions.

Primary and secondary amides of the type $RCONH_2$ and RCONHR react with nitrous acid in the same way as do the corresponding primary and secondary amines.

$$RCONH_2 + HONO \rightarrow RCOOH + N_2 + HOH$$
$$RCONHR + HONO \rightarrow RCON(NO)R + HOH$$

When diamides having their amide groups not far apart are heated, they lose ammonia to yield imides. See also **Imides.**

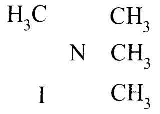
AMINATION. The process of introducing the amino group ($—NH_2$) into an organic compound is termed *amination.* An example is the reduction of aniline, $C_6H_5{\cdot}NH_2$, from nitrobenzene, $C_6H_5{\cdot}NO_2$. The reduction may be accomplished with iron and HCl. Only about 2% of the calculated amount of acid (to produce H_2 by reaction with iron) is required because of the fact that H_2O plus iron in the presence of ferrous chloride solution (ferrous and chloride ions) functions as the primary reducing agent. Such groups as nitroso (—NO), hydroxylamine (—NH·NH—), and azo (—N:N—) also yield amines by reduction. Amination also may be effected by the use of NH_3, in a process sometimes referred to as *ammonolysis.* An example is the production of aniline from chlorobenzene:

$$C_6H_5Cl + NH_3 \rightarrow C_6H_5{\cdot}NH_2 + HCl.$$

The reaction proceeds only under high pressure. In the ammonolysis of benzenoid sulfonic acid derivatives, an oxidizing agent is added to prevent the formation of soluble reduction products, such as $NaNH_4SO_4$, which commonly form. Oxygen-function compounds also may be subjected to ammonolysis: (1) methanol plus aluminum phosphate catalyst yields mono-, di-, and trimethylamines; (2) β-naphthol plus sodium ammonium sulfite catalyst (Bucherer reaction) yields β-naphthylamine; (3) ethylene oxide yields mono-, di-, and triethanolamines; (4) glucose plus nickel catalyst yields glucamine; and (5) cyclohexanone plus nickel catalyst yields cyclohexylamine.

AMINES. An amine is a derivative of NH_3 in which there is a replacement for one or more of the H atoms of NH_3 by an alkyl group, such as $—CH_3$ (methyl) or $—C_2H_5$ (ethyl); or by an aryl group, such as $—C_6H_5$ (phenyl) or $—C_{10}H_7$ (naphthyl). Mixed amines contain at least one alkyl and one aryl group as exemplified by methylphenylamine $CH_3{\cdot}N(H){\cdot}C_6H_5$. When one, two, and three H atoms are thus replaced, the resulting amines are known as *primary*, *secondary*, and *tertiary*, respectively. Thus, methylamine, CH_3NH_2, is a primary amine; di-methylamine, $(CH_3)_2NH$, is a secondary amine; and trimethylamine, $(CH_3)_3N$, is a tertiary amine. Secondary amines sometimes are called *imines*; tertiary amines, *nitriles*.

Quaternary amines consist of four alkyl or aryl groups attached to an N atom and, therefore, may be considered substituted ammonium bases. Commonly, they are referred to in the trade as quaternary ammonium compounds. An example is tetramethyl ammonium iodide.



$$\begin{array}{ccc} H_3C & & CH_3 \\ & N & CH_3 \\ I & & CH_3 \end{array}$$

The amines and quaternary ammonium compounds, exhibiting such great versatility for forming substitution products, are very important starting and intermediate materials for industrial organic syntheses, both on a small scale for preparing rare compounds for use in research and on a tonnage basis for the preparation of resins, plastics, and other synthetics. Very important industrially are the ethanolamines which are excellent absorbents for certain materials. See also **Ethanolamines.** Hexamethylene tetramine is a high-tonnage product used in plastics production. See also **Hexamine.** Phenylamine (aniline), although not as important industrially as it was some years ago, still is produced in quantity. Melamine is produced on a large scale and is the base for a series of important resins. See also **Melamine.** There are numerous amines and quaternary ammonium compounds that are not well known because of their importance as intermediates rather than as final products. Examples along these lines may include acetonitrile and acrylonitrile. See also **Acrylonitrile.**

Primary amines react (1) with nitrous acid, yielding (a) with alkylamine, nitrogen gas plus alcohol, (b) with warm arylamine, nitrogen gas plus phenol (the amino-group of primary amines is displaced by the hydroxyl group to form alcohol or phenol), (c) with cold arylamine, diazonium compounds, (2) with acetyl chloride or benzoyl chloride, yielding substituted amides, thus, ethylamine plus acetyl chloride forms N-ethylacetamide, $C_2H_5NHOCCH_3$, (3) with benzene-sulfonyl chloride, $C_6H_5SO_2Cl$, yielding substituted benzene sulfonamides, thus, ethylamine forms N-ethylbenzenesulfonamide, $C_6H_5SO_2—NHC_2H_5$, soluble in sodium hydroxide, (4) with chloroform, $CHCl_3$ with a base, yielding isocyanides (5) with HNO_3 (concentrated), yielding nitramines, thus, ethylamine reacts to form ethylnitramine, $C_2H_5—NHNO_2$.

Secondary amines react (1) with nitrous acid, yielding nitrosamines, yellow oily liquids, volatile in steam, soluble in ether. The secondary amine may be recovered by heating the nitrosamine with concentrated HCl, or hydrazines may be formed by reduction of the nitrosamines, e.g., methylaniline from methylphenylnitrosamine, $CH_3(C_6H_5)NNO$, reduction yielding unsymmetrical methylphenylhydrazine, $CH_3(C_6H_5)NHNH_2$, (2) with acetyl or benzoyl chloride, yielding substituted amides, thus, diethylamine plus acetyl chloride to form N, N-diethylacetamide $(C_2H_5)_2—NOCCH_3$, (3) with benzene sulfonyl chloride, yielding substituted benzene sulfonamides, thus, diethylamine reacts to form N,N-diethylbenzenesulfonamide, $C_6H_5SO_2N(C_2H_5)_2$, insoluble in NaOH.

Tertiary amines do not react with nitrous acid, acetyl chloride, benzoyl chloride, benzenesulfonyl chloride, but react with alkyl halides to form quaternary ammonium halides, which are converted by silver hydroxide to quaternary ammonium hydroxides. Quaternary ammonium hydroxides upon heating yield (1) tertiary amine plus alcohol (or, for higher members, olefin plus water). Tertiary amines may also be formed (2) by alkylation of secondary amines, e.g., by dimethyl sulfate, (3) from amino acids by living organisms, e.g., decomposition of fish in the case of trimethylamine.

AMINO ACIDS. The scores of proteins which make up about one-half of the dry weight of the human body and that are so vital to life functions are made up of a number of amino acids in various combinations and configurations. The manner in which the complex protein structures are assembled from amino acids is described in the entry on **Protein.** For some users of this book, it may be helpful to scan that portion of the protein entry which deals with the chemical nature of proteins prior to considering the details of this immediate entry on amino acids.

Although the proteins resulting from amino acid assembly are ultimately among the most important chemicals in the animal body (as well as plants), the so-called infrastructure of the proteins is dependent upon the amino acid building blocks. Although there are many hundreds of amino acids, only about 20 of these are considered very important to living processes, of which six to ten are classified as essential. Another three or four may be classified as quasi-essential, and ten to twelve may be categorized as nonessential. As more is learned about the fundamentals, protein chemistry, the scientific importance attached to specific amino acids varies. Usually, as the learning process continues, the findings tend to increase the importance of specific amino acids. Actually, the words *essential* and *nonessential* are not very good choices for naming categories of amino acids. Generally, those amino acids which the human body cannot synthesize at all or at a rate commensurate with its needs are called *essential amino acids* (EAA). In other words, for the growth and maintenance of a normal healthy body, it is essential that these amino acids be ingested as part of the diet and in the necessary quantities. To illustrate some of the indefinite character of amino acid nomenclature, some authorities classify histidine as an essential amino acid; others do not. The fact is that histidine is essential for the normal growth of the human infant, but to date it is not regarded as essential for adults. By extension of the preceding explanation, the term nonessential is taken to mean those amino acids that are really synthesized in the body and hence need not be present in food intake. This classification of amino acids, although amenable to change as the results of new findings, has been quite convenient in planning the dietary needs of people as well as of farm animals, pets, and also in terms of those plants that are of economic importance. The classification has been particularly helpful in planning the specific nutritional content of food substances involved in various aid and related programs for the people in needy and underdeveloped areas of the world.

Food Fortification with Amino Acids. In a report of the World Health Organization, the following observation has been made: "To determine the quality of a protein, two factors have to be distinguished, namely, the proportion of essential to nonessential amino acids and, secondly, the relative amounts of the essential amino acids . . . . The best pattern of essential amino acids for meeting human requirements was that found in whole egg protein or human milk, and comparisons of protein quality should be made by reference to the essential amino acid patterns of either of these two proteins." The ratio of each essential amino acid to the total sum is given for hen's egg and human and cow's milk in Table 1.

In the human body, tyrosine and cysteine can be formed from phenylalanine and methionine, respectively. The reverse transformations do not occur. Human infants have an ability to synthesize arginine and histidine in their bodies, but the speed of the process is slow compared with requirements.

TABLE 1. REPRESENTATIVE ESSENTIAL AMINO ACID PATTERNS
* A/E RATIO (Milligrams per gram of total essential amino acids)

	Hen's Egg (Whole)	Human Milk	Cow's Milk
Total "aromatic" amino acids	195	226	197
Phenylalanine	(114)	(114)	(97)
Tyrosine	(81)	(112)	(100)
Leucine	172	184	196
Valine	141	147	137
Isoleucine	129	132	127
Lysine	125	128	155
Total "S"	107	87	65
Cystine	(46)	(43)	(17)
Methionine	(61)	(44)	(48)
Threonine	99	99	91
Tryptophan	31	34	28

SOURCE: World Health Organization; FAO Nutrition Meeting Report Series, No. 37, Geneva, 1965.

* A/E Ratio equals ten times percentage of single essential amino acid to the total essential amino acids contained.

Several essential amino acids have been shown to be the limiting factor of nutrition in plant proteins. In advanced countries, the ratio of vegetable proteins to animal proteins in foods is 1.4:1. In underdeveloped nations, the ratio is 3.5:1, which means that people in underdeveloped areas depend upon vegetable proteins. Among vegetable staple foods, wheat easily can be fortified. It is used as flour all over the world. L-Lysine hydrochloride (0.2%) is added to the flour. Wheat bread fortified with lysine is used in several areas of the world; in Japan it is supplied as a school ration.

The situation of fortification in rice is somewhat more complex. Before cooking, rice must be washed (polished) with water. In some countries, the cooking water is allowed to boil over or is discarded. This significant loss of fortified amino acids must be considered. L-Lysine hydrochloride (0.2%) and L-threonine (0.1%) are shaped like rice grain with other nutrients and enveloped in a film. The added materials must hold the initial shape and not dissolve out during boiling, but be easily freed of their coating in the digestive organs.

The amino acids are arranged in accordance with essentiality in Table 2. Each of the four amino acids at the start of the table are all limiting factors of various vegetable proteins. Chick feed usually is supplemented with fish meals, but where the latter is in limited supply, soybean meals are substituted. The demand for DL-methionine, limiting amino acid in soybean meals, is now increasing. When seed meals, such as corn and sorghum, are used as feeds for chickens or pigs, L-lysine hydrochloride must be added for fortification. Lysine production is increasing upward to the level of methionine.

Early Research and Isolation of Amino Acids. Because of such rapid studies made within the past few decades in biochemistry and nutrition, these sciences still have a challenging aura about them. But, it is interesting to note that the first two natural amino acids were isolated by Braconnot in 1820. As shown by Table 3, these two compounds were glycine and leucine. Bopp isolated tyrosine from casein in 1849. Additional amino acids were isolated during the 1880s, but the real thrust into research in this field commenced in the very late 1800s and early 1900s with the work of Emden, Fischer, Mörner, and Hopkins and Cole. It is interesting to observe that Emil Fischer (1852–1919), German chemist and pioneer in the fields of purines and polypeptides, isolated three of these important compounds, namely, proline from gelatin in 1901, valine from casein in 1901, and hydroxyproline from gelatin in 1902. As an understanding of the role of amino acids in protein formation and of the function of proteins in nutrition progressed, the pathway was prepared for further isolation of amino acids. For example, in 1907, a combined committee representing the American Society of Biological Chemists and the American Physiological Society, proposed a formal classification of proteins into three major categories: (1) simple proteins, (2) conjugated proteins, and (3) derived proteins. The last classification embraces all denatured proteins and hydrolytic products or protein breakdown and no longer is considered as a general class.

Very approximate annual worldwide production of amino acids, their current method of preparation (not exclusive), and general characteristics are given in Table 2.

The *isoelectric point* is very important in the preparation and separation of amino acids and proteins. Protein solubility varies markedly with pH and is at a minimum at the isoelectric point. By raising the salt concentration and adjusting pH to the isoelectric point, it is often possible to obtain a precipitate considerably enriched in the desired protein and to crystallize it from a heterogenous mixture.

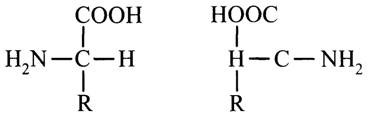
Chemical Nature of Amino Acids

In a very general way, an amino acid is any organic acid which incorporates one or more amino groups. This definition includes a multitude of substances of most diverse structure. There are seemingly limitless related compounds of differing molecular size and constitution which incorporate varying kinds and numbers of functional groups. Most extensive study has centered around the relatively small group of alpha-amino acids which are combined in amide linkage to form proteins. With few exceptions, these compounds possess the general structure NH_2CHRCO_2H, where the amino group occupies a position on the carbon atom *alpha* to that of the carboxyl group, and where the side chain R may be of diverse composition and structure.

Few products of natural origin are as versatile in their behavior and properties as are the amino acids, and few have such a variety of biological duties to perform. Among their general characteristics would be included:

(a) Water-soluble and amphoteric electrolytes, with the ability to form acid salts and basic salts and thus act as buffers over at least two ranges of pH (hydrogen ion concentration).
(b) Dipolar ions of high electric moment with a considerable capacity to increase the dielectric constant of the medium in which they are dissolved.
(c) Compounds with reactive groups capable of a wide range of chemical alterations leading readily to a large variety of degradation, synthetic, and transformation products, such as esters, amides, amines, anhydrides, polymers, polypeptides, diketopiperazines, hydroxy acids, halogenated acids, keto acids, acylated acids, mercaptans, shorter-or longer-chained acids, and pyrrolidine and piperidine ring forms.
(d) Indispensable components of the diet of all animals including humans.
(e) Participants in crucial metabolic reactions on which life depends, and substrates for a variety of specific enzymes in vitro.
(f) Binders of metals of many kinds.

Optical Properties. With the exception of glycine ($NH_2CH_2CO_2H$) and amino-malonic acid [$NH_2CH(CO_2H)_2$], all α-amino acids which are classifiable according to the general formula previously given exist in at least two different optically isometric forms. The optical isomers of a given amino acid possess identical empirical and structural formulas, and are indistinguishable from each other on the basis of their chemical and physical properties, with the singular exception of their effect on plane polarized light. This may be illustrated with the two optically active forms as shown by



$$\begin{array}{c} \text{COOH} \\ | \\ \text{H}_2\text{N}-\text{C}-\text{H} \\ | \\ \text{R} \end{array} \qquad \begin{array}{c} \text{HOOC} \\ | \\ \text{H}-\text{C}-\text{NH}_2 \\ | \\ \text{R} \end{array}$$

L-form D-form

mirror
image

One form (L-form) exhibits the ability to rotate the plane of polarization of plane polarized light to the left (levorotatory), whereas the other form (D-form) rotates the plane to the right (dextrorotatory). Although the direction of optical rotation exhibited by these optically active forms is different, the magnitude of their respective rotations is the same. If equal amounts of *dextro* and *levo* forms are admixed, the optical effect of each isomer is neutralized by the other, and an optically inactive product known as a *racemic modification* or *racemate* is secured.

The ability of the alanine molecule, for example, to exist in two stereoisomeric forms can be attributed to the fact that the α-carbon atom of this compound is attached to four different groups which may vary in their three-dimensional spatial arrangement. Compounds of this type do not possess complete symmetry when viewed from a purely geometrical standpoint and hence are generally referred to as *asymmetric*. As a consequence of this molecular asymmetry, the four covalent bonds of an asymmetric carbon atom can be aligned in a manner such that a regular tetrahedron is formed by the straight lines connecting their ends. Hence, two different tetrahedral arrangements of the groups about the asymmetric carbon atom can be devised so that these structures relate to one another as an object relates to its mirror image, or as the right hand relates to the left hand. Molecules of this type are endowed with the property of optical activity and, together with their nonsuperimposable mirror images, are generally referred to as *enantiomorphs, enantiomers, antimers*, or *optical antipodes*.

Classification. In accordance with the structure of the R-group, the amino acids of primary importance can be classified into eight groups. Additional amino acids composing protein are not included in this classification, because they occur infrequently. See Table 4.

Normally, amino acids exist as dipolar ions. $RCH(NH_3^+)COO^-$, in a neutral state, where both amino and carboxyl groups are ionized. The

TABLE 2. IMPORTANT NATURAL AMINO ACIDS AND PRODUCTION

Amino Acid	World Annual Production, tons	Present Mode of Manufacture	Characteristics
		Essential Amino Acids	
DL-Methionine	10^4	Synthesis from acrolein and mercaptan	First limiting amino acid for soybean
L-Lysine. HCl	10^3	Fermentation (AM)*	First limiting amino acid for all cereals
L-Threonine	10	Fermentation (AM)	Second limiting amino acid for rice
L-Tryptophan	10	Synthesis from acrylonitrile and resolution	Second limiting amino acid for corn
L-Phenylalanine	10	Synthesis from phenylacetaldehyde and resolution	Rich in plant protein
L-Valine	10	Fermentation (AM)	
L-Leucine	10	Extraction from protein	
L-Isoleucine	10	Fermentation (WS)**	Deficient in some cases
		Quasi-Essential Amino Acids	
L-Arginine. HCl	10^2	Synthesis from *L*-ornithine Fermentation (AM)	Essential to human infants
L-Histidine. HCl	10	Extraction from protein	
L-Tyrosine	10	Enzymation of phenol and serine	Limited substitute for phenylalanine
L-Cysteine L-Cystine	10	Extraction from human hair	Limited substitute for methionine
		Nonessential Amino Acids	
L-Glutamic acid	10^5	Fermentation (WS) Synthesis from acrylonitrile and resolution	MSG, taste enhancer
Glycine	10^3	Synthesis from formaldehyde	Sweetener
DL-Alanine	10^2	Synthesis from acetaldehyde	
L-Aspartic acid	10^2	Enzymation of fumaric acid	Hygienic drug
L-Glutamine	10^2	Fermentation (WS)	Anti-gastroduodenal ulcer drug
L-Serine	<10	Synthesis from glycolonitrile and resolution	Rich in raw silk
L-Proline	<10	Fermentation (AM)	Rich in gelatin
L-Hydroxyproline	<10	Extraction from gelatin	
L-Asparagine	<10	Synthesis from L-aspartic acid	Neurotropic metabolic regulator
L-Alanine	<10	Enzymation of L-aspartic acid	Rich in degummed white silk
L-Dihydroxyphenylalanine	10^2	Synthesis from piperonal, vanillin, or acrylonitrile and resolution	Specific drug for Parkinson's disease
L-Citrulline	<10	Fermentation (AM)	Ammonia detoxicant
L-Ornithine	<10	Fermentation (AM)	

*AM, artificial mutant; **WS, wild strain. MSG, monosodium glutamate.

TABLE 3. FIRST ISOLATION OF AMINO ACIDS

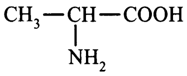
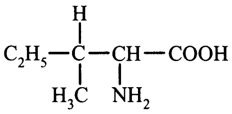
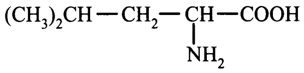
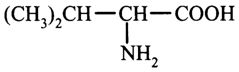
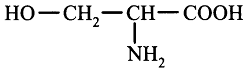
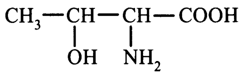
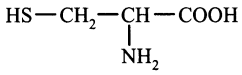
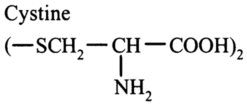
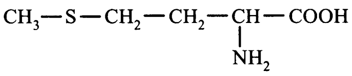
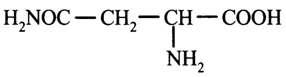
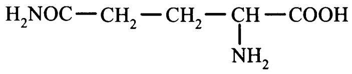
Abbreviation	Name and Formula	First Isolation and (Source)	Isoelectric Point
	Neutral Amino Acids—Aliphatic Type		
Ala	Alanine  $CH_3-\underset{\displaystyle NH_2}{\underset{\vert}{CH}}-COOH$	1879 by Schutzenberger 1888 by Weyl (silk fibroin)	6.0
Gly	Glycine NH_2-CH_2-COOH	1820 by Braconnot (gelatin)	6.0
Ile	Isoleucine  $C_2H_5-\underset{\displaystyle H_3C}{\underset{\vert}{\overset{\displaystyle H}{\overset{\vert}{C}}}}-\underset{\displaystyle NH_2}{\underset{\vert}{CH}}-COOH$	1904 by Ehrlich (fibrin)	6.0
Leu	Leucine  $(CH_3)_2CH-CH_2-\underset{\displaystyle NH_2}{\underset{\vert}{CH}}-COOH$	1820 by Braconnot (muscle fiber; wool)	6.0
Val	Valine  $(CH_3)_2CH-\underset{\displaystyle NH_2}{\underset{\vert}{CH}}-COOH$	1901 by Fischer (casein)	6.0
	Neutral Amino Acids—Hydroxy Type		
Ser	Serine  $HO-CH_2-\underset{\displaystyle NH_2}{\underset{\vert}{CH}}-COOH$	1865 by Cramer (sericine)	5.7
Thr	Threonine  $CH_3-\underset{\displaystyle OH}{\underset{\vert}{CH}}-\underset{\displaystyle NH_2}{\underset{\vert}{CH}}-COOH$	1925 by Gortner and Hoffman 1925 by Schryver and Buston (oat protein)	6.2
	Neutral Amino Acids—Sulfur-Containing Type		
Cys	Cysteine  $HS-CH_2-\underset{\displaystyle NH_2}{\underset{\vert}{CH}}-COOH$	----------------------	5.1
Cys Cys	Cystine  $(-SCH_2-\underset{\displaystyle NH_2}{\underset{\vert}{CH}}-COOH)_2$	1899 by Mörner (horn) 1899 by Emden	4.6
Met	Methionine  $CH_3-S-CH_2-CH_2-\underset{\displaystyle NH_2}{\underset{\vert}{CH}}-COOH$	1922 by Mueller (casein)	5.7
	Neutral Amino Acids—Amide Type		
Asn	Asparagine  $H_2NOC-CH_2-\underset{\displaystyle NH_2}{\underset{\vert}{CH}}-COOH$	1932 by Damodaran (edestin)	5.4
Gln	Glutamine  $H_2NOC-CH_2-CH_2-\underset{\displaystyle NH_2}{\underset{\vert}{CH}}-COOH$	1932 by Damodaran, Jaaback, and Chibnall (gliadin)	5.7

TABLE 3. (*continued*)

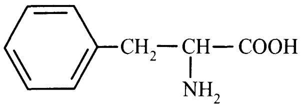
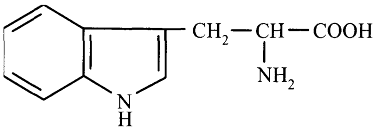
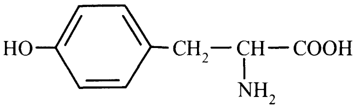
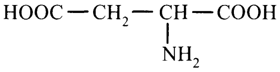
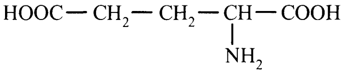
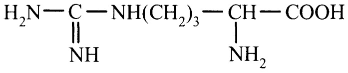
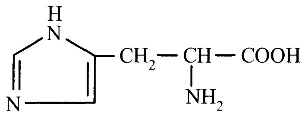
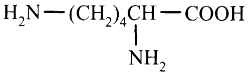
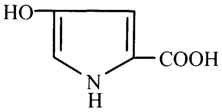
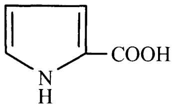
Abbreviation	Name and Formula	First Isolation and (Source)	Isoelectric Point
	Neutral Amino Acids—Aromatic Type		
Phe	Phenylalanine 	1881 by Schulze and Barbieri (lupine seedings)	5.5
Trp	Tryptophan 	1902 by Hopkins and Cole (casein)	5.9
Tyr	Tyrosine 	1849 by Bopp (casein)	5.7
	Acidic Amino Acids		
Asp	Aspartic Acid  $HOOC-CH_2-\underset{NH_2}{CH}-COOH$	1868 by Ritthausen (conglutin; legumin)	2.8
Glu	Glutamic acid  $HOOC-CH_2-CH_2-\underset{NH_2}{CH}-COOH$	1866 by Ritthausen (gluten-fibrin)	3.2
	Basic Amino Acids		
Arg	Arginine  $H_2N-\underset{NH}{\underset{\|}{C}}-NH(CH_2)_3-\underset{NH_2}{CH}-COOH$	1895 by Hedin (horn)	11.2
His	Histidine 	1896 by Kossel (sturine) 1896 by Hedin (various protein hydrolysates)	7.6
Lys	Lysine  $H_2N-(CH_2)_4\underset{NH_2}{CH}-COOH$	1889 by Dreschel (casein)	9.7
	Imino Acids		
Hyp	Hydroxyproline 	1902 by Fischer (gelatin)	5.8
Pro	Proline 	1901 by Fischer (casein)	6.3

TABLE 4. STRUCTURAL CLASSIFICATION OF AMINO ACIDS

NEUTRAL AMINO ACIDS		
Alphatic-type	*Hydroxy-type*	*Sulfur-containing*
Glycine	Serine	Cysteine
Alanine	Threonine	Cystine
Valine		Methionine
Leucine		
Isoleucine		
Amide-type	*Aromatic-type*	
Asparagine	Phenylalanine	
Glutamine	Tryptophan	
	Tyrosine	
ACIDIC AMINO ACIDS		
Aspartic acid		
Glutamic acid		
BASIC AMINO ACIDS		
Histidine		
Lysine		
Arginine		
IMINO ACIDS		
Proline		
Hydroxyproline		

dipolar form, $RCH(NH_2)COOH$ may be considered, but the dipolar form predominates for the usual monoamino monocarboxylic acid and it is estimated that these forms occur 10^5 to 10^6 times more frequently than the nonpolar forms. Amino acids decompose thermally at what might be considered a relatively high temperature (200–300°C). The compounds are practically insoluble in organic solvents, have low vapor pressure, and do not exhibit a precisely defined melting point.

The ionic states of a simple α-amino acid are given by

$$\underset{\text{(Cationic form; acidic)}}{RCH(NH_3^+)COOH} \underset{+H^+}{\overset{-H+(K_1)}{\rightleftharpoons}}$$

$$\underset{\text{(Dipolar form; neutral)}}{RCH(NH_3^+)COO^-} \underset{+H^+}{\overset{-H+(K_2)}{\rightleftharpoons}} \underset{\text{(Anionic form; basic)}}{RCH(NH_2)COO^-}$$

In accordance with the change of the ionic state, dissociation constants are

$$K_1(COOH) = \frac{[H^+][RCH(NH_3^+)COO^-]}{[RCH(NH_3^+)COOH]}$$

$$K_2(NH_3^+) = \frac{[H^+][RCH(NH_2)COO^-]}{[RCH(NH_3^+)COO^-]}$$

Inasmuch as $pK = -\log K$, the values for glycine are $pK_1 = 2.34$ and $pK_2 = 9.60$ (in aqueous solution at 25°C). The homologous amino acids indicate similar values. The pH at which acidic ionization balances basic ionization is termed the *isoelectric point* (pH_I), corresponding to

$$[RCH(NH_3^+)COOH] = [RCH(NH_2)COO^-]$$

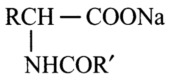
Thus, from these formulas, the pH_I is

$$pH_I = \tfrac{1}{2}(pK_1 + pK_2)$$

Formation of Salts. Amino acids have certain characteristics of both organic bases and organic acids because they are amphoteric. As amines, the amino acids form stable salts, such as hydrochlorides or aromatic sulfonic acid salts. These are used as selective precipitants of certain amino acids. As organic acids, the amino acids form complex salts with heavy metals, the less soluble salt being used for amino acid separation.

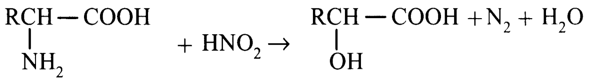
Esters. When heated with the equivalent amount of a strong acid, usually hydrochloric acid in absolute alcohol, amino acids form esters. These are obtained as hydrochlorides.

Acylation. In alkaline solution, amino acids react with acid chlorides or acid anhydrides to form acyl compounds of the type



$$\underset{\substack{|\\ NHCOR'}}{RCH}-COONa$$

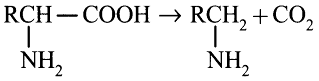
Van Slyke Reaction (Deamination). With excess nitrous acid, -amino acids react to form -hydroxyl acids on a quantitative basis. Nitrogen gas is generated.



$$\underset{\substack{|\\ NH_2}}{RCH}-COOH + HNO_2 \rightarrow \underset{\substack{|\\ OH}}{RCH}-COOH + N_2 + H_2O$$

The reaction is completed within five minutes at room temperature. Thus, measurement of the volume of nitrogen generated can be used in amino acid determinations.

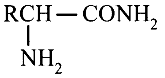
Decarboxylation. When heated with inert solvents, such as kerosene, amino acids form amines



$$\underset{\substack{|\\ NH_2}}{RCH}-COOH \rightarrow \underset{\substack{|\\ NH_2}}{RCH_2} + CO_2$$

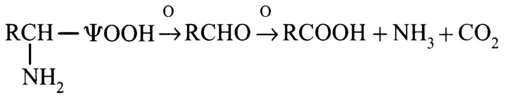
Decarboxylative enzymes may react specifically with amino acids having free polar groups at the ω position. Cadaverine can be produced from lysine, histamine from histidine, and tyramine from tyrosine.

Formation of Amides. When condensed with ammonia or amines, amino acid esters form acid amides:



$$\underset{\substack{|\\ NH_2}}{RCH}-CONH_2$$

Oxidation. Oxidizing agents easily decompose α-amino acids, forming the corresponding fatty acid with one less carbon number:



$$\underset{\substack{|\\ NH_2}}{RCH}-\Psi OOH \xrightarrow{O} RCHO \xrightarrow{O} RCOOH + NH_3 + CO_2$$

Ninhydrin Reaction. A neutral solution of an amino acid will react with ninhydrin (triketohydrindene hydrate) by heating to cause oxidative decarboxylation. The central carbonyl of the triketone is reduced to an alcohol. This alcohol further reacts with ammonia formed from the amino acid and causes a red-purplish color. Since the reaction is quantitative, measurement of the optical density of the color produced is an indication of amino acid concentration. Imino acids, such as hydroxyproline and proline, develop a yellow color in the same type of reaction.

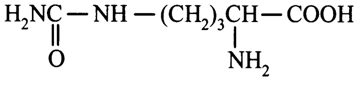
Maillard Reaction. In amino acids, the amino group tends to form condensation products with aldehydes. This reaction is regarded as the cause of the browning reaction when an amino acid and a sugar coexist. A characteristic flavor, useful in food preparations, is evolved along with the color in this reaction.

Ion-exchange Separations. Because amino acids are amphoteric, they behave as acids or bases, depending upon the pH of the solution. This makes it possible to adsorb amino acids dissolved in water on either a strong-acid cation exchange resin; or a strong-base anion exchange resin. The affinity varies with the amino acid and the solution pH. Ion-exchange resins are widely used in amino acid separations.

Production of Amino Acids. There are three means available for making (or separating) amino acids in large quantity lots: (1) *extraction* from natural protein; (2) *fermentation*; and (3) chemical *synthesis*. During the early investigations of amino acids, the first method was widely used and still applies to four amino acids. See Table 3.

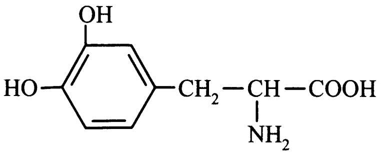
L-Leucine is easily extracted in quantity from almost any type of vegetable protein hydrolyzates. Cystine is extracted from the human-hair hydrolyzate. L-Histidine is obtainable from the blood of animals, but future yields may stem from fermentation inasmuch as some artificial mutants of bacteria have been discovered. Gelatin is the prime source of L-hydroxyproline.

Natural amino acids, normally not contained in proteins, but which are effective in medicine, include citrulline, ornithine, and dihydroxyphenylalanine. These are not listed in Tables 2 and 3. Citrulline (Cit) with an isoelectric point of 5.9 was isolated by Koga in 1914; by Odake in 1914; and by Wada in 1930. It has the formula

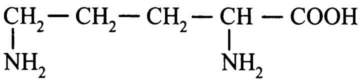


$$\underset{\overset{\|}{O}}{H_2NC} - NH - (CH_2)_3\underset{\overset{|}{NH_2}}{CH} - COOH$$

Dihydroxyphenylalanine (Dopa) with an isoelectric point of 5.5 was isolated by Torquati in 1913; and by Guggenheim in 1913. It has the formula



Ornithine (Orn) with an isoelectric point of 9.7 was isolated by Riesser in 1906 from arginine. It has the formula



$$\underset{\overset{|}{NH_2}}{CH_2} - CH_2 - CH_2 - \underset{\overset{|}{NH_2}}{CH} - COOH$$

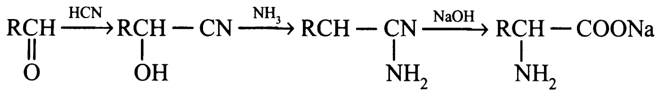
Fermentation Methods. Numerous microorganisms can synthesize the amino acids required to support their life from a simple carbon source and an inorganic nitrogen source, such as ammonium or nitrate salts, or nitrogen gas.

Japanese microbiologists, in 1956, first succeeded in developing industrial production of L-glutamic acid by a microbiological process. As of the present, nearly all common amino acids can be produced on a low cost industrial scale by fermentation. From microbiological studies, it has been ascertained that some microbial stains isolated from natural sources serve to excrete and accumulate a large amount of a particular amino acid in the cultural broth under carefully controlled conditions. The production of glutamic acid is produced by adding a selected bacterial strain and culturing aerobically for one to two days in a chemically defined medium which contains carbon sources, such as sugar or acetate, and nitrogen sources, such as ammonium salts. About 50% (wt) of the carbon sources can be converted to glutamate.

Genetic techniques have been used to improve the ability of microorganisms to accumulate amino acids. Several amino acids are manufactured from their direct precursors by the use of microbially produced enzymes. For example, bacterial L-aspartate β-carboxylase is used for the production of L-alanine from L-aspartic acid.

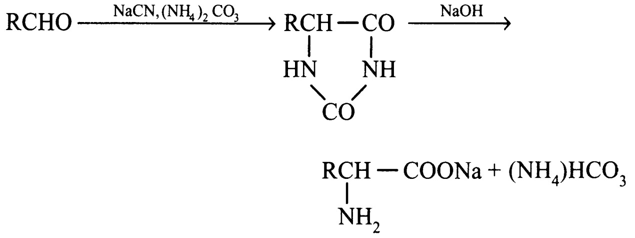
In isolating the amino acids from the fermentation broth, chromatographic separations using ion-exchange resins are the most important commercial method. Precipitation with compounds which yield insoluble salts with amino acids are also used. Purification is possible by crystallization through careful adjustment of the isoelectric point, at which point the amino acid is least soluble.

There are several laboratory-size methods for synthesizing amino acids, but few of these have been scaled up for industrial production. Glycine and DL-alanine are made by the Strecker synthesis, commencing with formaldehyde and acetaldehyde, respectively. In the Strecker synthesis, aldehydes react with hydrogen cyanide and excess ammonia to give amino nitriles which, in turn, are converted into α-amino acids upon hydrolysis.



$$\underset{\overset{\|}{O}}{RCH} \xrightarrow{HCN} \underset{\overset{|}{OH}}{RCH} - CN \xrightarrow{NH_3} \underset{\overset{|}{NH_2}}{RCH} - CN \xrightarrow{NaOH} \underset{\overset{|}{NH_2}}{RCH} - COONa$$

The Hydantoin Process. Hydantoins are produced by reacting aldehydes with sodium cyanide and ammonium carbonate. Upon hydrolysis, α-amino acids will be yielded



The production of α-amino acids by chemical synthesis yields a mixture of DL forms. The D-form of glutamic acid has no flavor-enhancing properties and thus requires transformation into the optically active form insofar as monosodium glutamate is concerned. The three methods for separating the optical isomers are: (1) preferential inoculation method; (2) the diasteroisomer method; and (3) the acylase method.

Hauromi Oeda, Ajinomoto Co., Inc., Kawasaki, Japan.

AMINO RESINS. A family of resins resulting from an addition reaction between formaldehyde and compounds, such as aniline, ethylene urea, dicyandiamide, melamine, sulfonamide, and urea. The resins are thermosetting and have been used for many years in such products as textile-treating agents, laminating coatings, wet-strength paper coatings, and wood adhesives. The urea and melamine compounds are the most widely used. Both of these basic resins are water white (transparent). However, the resins readily accept pigments and opacifying agents. The addition of cellulose filler can be used to reduce light transmission. Where color is unimportant, various materials are added to the melamine resin compounds, including macerated fabric, glass fiber, and wood flour. Wood flour frequently is added to the urea resins to yield a low cost industrial material.

Advantages claimed for amino resins include: (1) good electrical insulation characteristics, (2) no transfer of tastes and odors to foods, (3) self-extinguishing burning characteristics, (4) resistance to attack by oils, greases, weak alkalis and acids, and organic solvents, (5) abrasion resistance, (6) good rigidity, (7) easy fabrication by economical molding procedures, (8) excellent resistance to deformation under load, (9) good subzero characteristics with no tendency to become brittle, and (10) marked hardness.

Amino resins are fabricated principally by transfer and compression molding. Injection molding and extrusion are used on a limited scale. Urea resins are not recommended for outdoor exposure. The resins show rather high mold shrinkage and some shrinkage with age. The melamines are superior to the ureas insofar as resistance to heat and boiling water, acids, and alkalis is concerned.

Some of the hundreds of applications for amino resins include: closures for glass, metal, and plastic containers; electrical wiring devices; appliance knobs, dials, handles, and push buttons; lamp shades and lighting diffusers; organ and piano keys; dinnerware; food service trays; food-mixer housings; switch parts; decorative buttons; meter blocks; aircraft ignition parts; heavy duty switch gear; connectors; and terminal strips. Not all of the urea or melamine amino resins are suited to all of the foregoing uses. Because of the large number of fillers and additives available, the overall range of use of this family of resins is large.

AMITOSIS. Cell division without mitosis; a splitting of the cell into two parts without the previous duplication of chromosomes and segregation of the duplicated chromosomes into two separate groups. Formerly amitosis was thought to be the method of cell division for many of the simpler, one-celled forms of life. Improved techniques of microscopic study, however, have shown that there is some form of mitosis even in these simple forms. Today amitosis is recognized as a rare and abnormal form of cell division which produces cells with a limited survival.

AMMINES. Dry ammonia gas reacts with dehydrated salts of some of the metals to form solid ammines. Ammines, upon warming, evolve ammonia, sometimes with final decomposition of the salt itself, in a manner analogous to the decomposition of certain hydrates. The ammines of chromium(III)(Cr^{3+}), cobalt(III), platinum(IV), and other

$[Co(NH_3)_6]Cl_3$
410

$[Co(NH_3)_5(NO_2)]Cl_2$
240

$[Co(NH_3)_4(NO_2)_2]Cl$
95

$[Co(NH_3)_3(NO_2)_3]$
1.5

$K[Co(NH_3)_2(NO_2)_4]$
95

$K_2[Co(NH_3)(NO_2)_5]$
240

$K_3[Co(NO_2)_6]$
420

(a)

$[Cr(NH_3)_6]X_3$
$[Cr(NH_3)_5(H_2O)]X_3$
$[Cr(NH_3)_4(H_2O)_2]X_3$
$[Cr(NH_3)_3(H_2O)_3]X_3$
$[Cr(NH_3)_2(H_2O)_4]X_3$

X = unit anion

(b)

Two series of ammines: (a) Square bracket contains the ion. The equivalent electrical conductivity is shown below each compound. The number of neutral groups, e.g., (NH_3), on metal, e.g., Co, is varied from 6 to 0. (b) The number of neutral groups is constant, but the groups are varied.

metals have been studied in detail. Two series of ammines are shown in the accompanying diagram, the first being one in which the neutral ammonia group is replaced step by step by the negative nitro group (NO_2^-), and the second, one in which the neutral ammonia group is replaced step by step by the neutral H_2O group.

The neutral group of the complex may be replaced step by step by the following negative groups: Cl^-, Br^-, I^-, F^-, OH^-, NO_2^-, NO_3^-, CN^-, CNS^-, SO_4^{2-}, CO_3^{2-}, $C_2O_4^{2-}$; or by the following neutral groups: H_2O, NO, NO_2, SO_2, S, N_2H_4, H_2NOH, CO, C_2H_5OH, C_6H_6. All neutral groups are of substances capable of independent existence.

In the ammines, trivalent metals, such as cobalt(III) and chromium(III) and iron(III), possess a coordination number of 6, this number being the sum of the unit replacements on the metal in the complex ion. Since a regular octahedron has six corners equidistant from the center, it is assumed that the metal occupies the center and each of the six replacing groups occupies a corner of a regular octahedron. Support for this assumption is offered by the x-ray examination of these ammines. When there is only one of the six groups replaced by a second group, as in $[Co(NH_3)_5(NO_2)]Cl_2$, and in $[Cr(NH_3)_5(H_2O)]X_3$ the octahedral placement of groups supplies only one form, but when two of the six groups are replaced by a second group, as in $[Co(NH_3)_4(NO_2)_2]Cl$, and in $[Cr(NH_3)_4(H_2O)_2]X_3$, two different octahedral corner arrangements are possible depending upon whether the two replacing groups are adjacent (cis-form) or opposite (trans-form). Two substances differing in physical properties and corresponding to these two forms are known. Further, when three divalent groups, e.g., $3C_2O_4^{2}A$ are present in the complex, two arrangements—not identical but mirror-images of each other—are possible. Two optically active substances are known in such cases corresponding to these two stereoisomeric forms.

Six is the ordinary coordination number for metallic ammines and similar complexes. Additional examples are $K_2[Pt(NH_3)_2(CN)_4]$, $[Ni(NH_3)_6]Cl_2$, $K_4[Fe(CN)_6]$, $K_3[Fe(CN)_6]$, $K_2[Fe(CN)_5(NO)]$, $K_2[SiF_6]$, $[Ca(NH_3)_6]Cl_2$. But, for the elements boron, carbon, and nitrogen four is the coordination number, e.g., $[BH_4]Cl$, $[CH_4]$, $[NH_4]Cl$, and in these substances the groups are assumed to occupy the corners of a regular tetrahedron; in $K_4[Mo(CN)_8]$ and $[Ba(NH_3)_8]Cl_2$ the coordination number is eight, and the groups are assumed to occupy the corners of a cube.

AMMETER. See **Electrical Instruments.**

AMMONIA. Known since ancient times, ammonia, NH_3, has been commercially important for well over 100 years and has become the second largest chemical in terms of tonnage and the first chemical in value of production. The first practical plant of any magnitude was built in 1913. Worldwide production of NH_3 as of the early 1980s is estimated at 100 million metric tons per year or more, with the United States accounting for about 14% of the total production. A little over three-fourths of ammonia production in the United States is used for fertilizer, of which nearly one-third is for direct application. An estimated 5.5% of ammonia production is based in the manufacture of fibers and plastics intermediates.

Properties. At standard temperature and pressure, NH_3 is a colorless gas with a penetrating, pungent-sharp odor in small concentrations which, in heavy concentrations, produces a smothering sensation when inhaled. Formula weight is 17.03, mp −77.7°C, bp −33.35°C, and sp gr 0.817 (at −79°C) and 0.617 (at 15°C). Ammonia is very soluble in water, a saturated solution containing approximately 45% NH_3 (weight) at the freezing temperature of the solution and about 30% (weight) at standard conditions. Ammonia dissolved in water forms a strongly alkaline solution of ammonium hydroxide, NH_4OH. The univalent radical NH_4^+ behaves in many respects like K^+ and Na^+ in vigorously reacting with acids to form salts. Ammonia is an excellent nonaqueous electrolytic solvent, its ionizing power approaching that of water. Ammonia burns with a greenish-yellow flame.

Ammonia derives its name from sal ammoniac, NH_4Cl, the latter material having been produced at the Temple of Jupiter Ammon (Libya) by distilling camel dung. During the Middle Ages, NH_3 was referred to as the spirits of hartshorn because it was produced by heating the hoofs and horns of oxen. The composition of ammonia was first established by Claude Louis Berthollet (France, ca. 1777). The first significant commercial source of NH_3 (during the 1880s) was its production as a by-product in the making of manufactured gas through the destructive distillation of coal. See also **Coal Tar and Derivatives.**

Nitrogen fixation is a term assigned to the process of converting nitrogen in the air to nitrogen compounds. Although some bacteria in soil are capable of this process, N_2 as an ingredient of fertilizer is required for soils that are depleted by crop production. The production of synthetic NH_3 is the most important industrial nitrogen-fixation process. See also **Fertilizers.**

Synthesis of Ammonia

The first breakthrough in the large-scale synthesis of ammonia resulted from the work of Fritz Haber (Germany, 1913), who found that ammonia could be produced by the direct combination of two elements, nitrogen and hydrogen, ($N_2 + 3H_2 \rightleftharpoons 2NH_3$) in the presence of a catalyst (iron oxide with small quantities of cerium and chromium) at a relatively high temperature (550°C) and under a pressure of about 200 atmospheres, representing difficult processing conditions for that era. Largely because of the urgent requirements for ammonia in the manufacture of explosives during World War I, the process was adapted for industrial-quality production by Karl Bosch, who received one-half of the 1931 Nobel Prize for chemistry in recognition of these achievements. Thereafter, many improved ammonia-synthesis systems, based on the Haber-Bosch process, were commercialized, using various operating conditions and synthesis-loop designs.

The principal features of an NH_3 synthesis process system are the converter designs, operating conditions, method of product recovery, and type of recirculation equipment. Most current systems operate at or above the pressure used in the original Haber-Bosch process. Converter designs have either a single continuous catalyst bed, which may or may not have heat-exchange cooling for controlling reaction heat, or several catalyst beds with provision for temperature control between the beds.

Claude Process. The original Claude process was one of the first systems to use a high operating pressure (1000 atmospheres), achieving 40% conversion without recycling. This system used multiple converters in a series-parallel arrangement. The present Claude process[1] operates at 340–650 atmospheres, using a single converter with continuous catalyst-charged tubes externally cooled to remove the heat of reaction. Approximate hydrogen conversion is 30–34 mole percent per pass. The pressure is increased gradually to compensate for catalyst aging and loss in activity. Product recovery is by simple condensation in a water-cooled condenser. Unreacted gas is recycled by compressor.

Casale Process. This is another high-pressure conversion system, using synthesis pressures of 450–600 atmospheres, which also permits

---

[1] Developed by Grande Pariosse and L'Air Liquide.

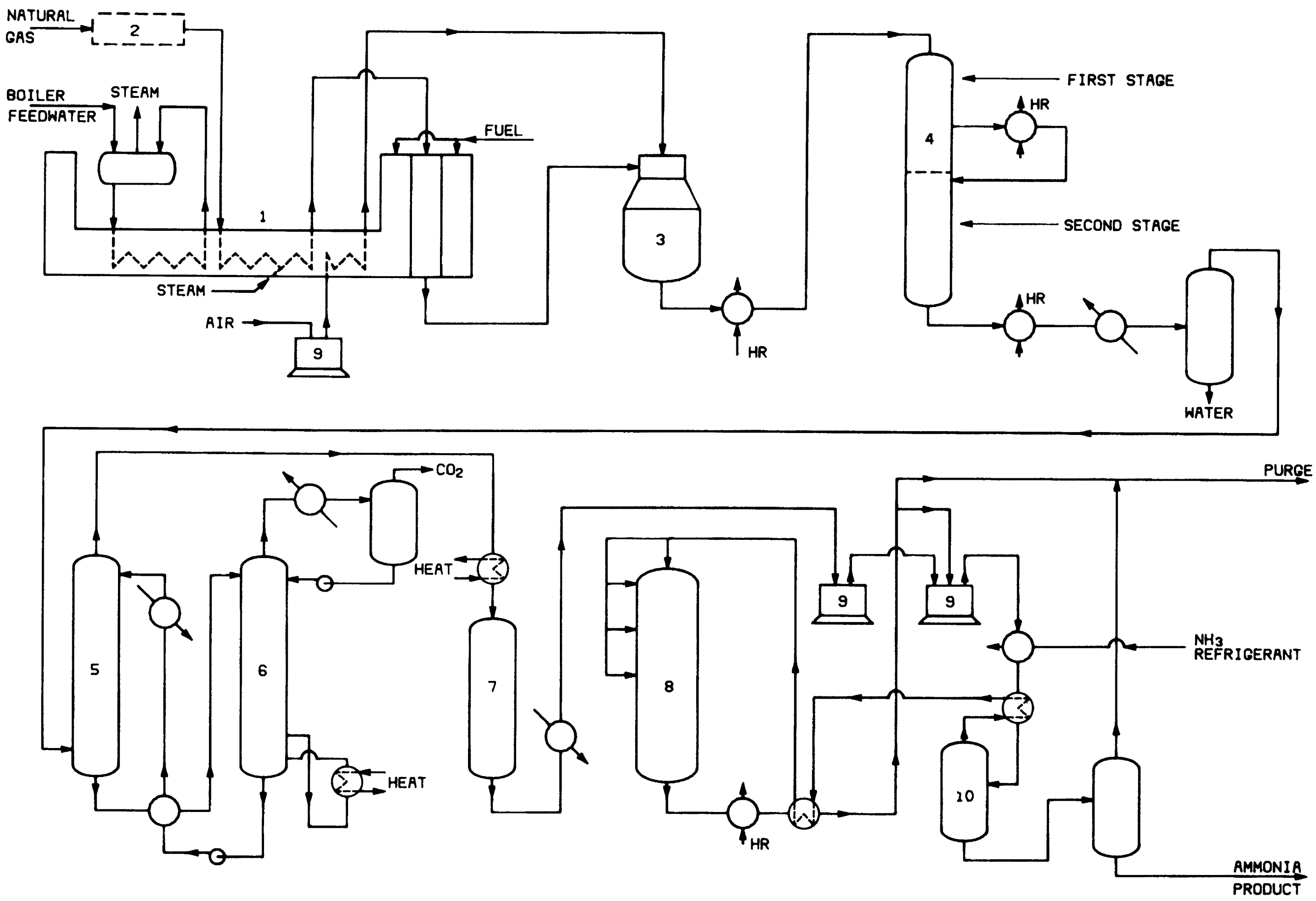


Fig. 1. Ammonia production process: (1) Primary reformer, (2) desulforization, (3) secondary reformer, (4) CO shift converter (in two stages), (5) CO_2 absorber, (6) CO_2 stripper, (7) methanator, (8) NH_3 converter, (9) compressor, (10) separator. HR = heat recovery. (*M. W. Kellogg.*)

hydrogen conversions in the 30 mole percent range. As in the Claude process, the high pressure allows NH_3 to be recovered from the converter effluent by water cooling. The Casale converter uses a single catalyst bed with internal heat-exchange surfaces. Reaction rate and temperature rise across the catalyst are controlled by the internal exchanger and retaining 2–3 mole percent NH_3 in the converter feed. An ejector is used to remove unreacted gas. This eliminates the need for a mechanical recycle compressor, but requires high feed-gas pressures to supply the energy required for the ejector.

Low-Pressure Processes. Several systems use low synthesis pressures with hydrogen conversion below 30 mole percent and product recovery by water and refrigeration.

Synthesis-Gas-Production Processes. These processes were improved and developed as a result of changes in feedstock availability and economics. Before World War II, most NH_3 plants obtained H_2 by reacting coal or coke with steam in the water-gas process. A small number of plants used water electrolysis or coke-oven by-product hydrogen. The subsequent low-cost availability of natural gas brought about steam-hydrocarbon reforming as the major source of H_2 for the NH_3 synthesis gas.

Partial oxidation processes to produce H_2 from natural gas and liquid hydrocarbons were also developed after World War II and accounted for 15% of the synthetic NH_3 capacity by 1962. The steam-hydrocarbon reforming process[2] was developed in 1930. In this process, methane was mixed with an excess of steam at atmospheric pressure, and the mixture reformed inside nickel-catalyst-filled alloy furnace tubes. The heat of reaction was supplied by externally heating the catalyst-filled tubes to about 871°C. Since the late 1950s, improvements in the tubular-reforming technology and metallurgy have brought about the utilization of high-pressure ($>$ 24 atmospheres) reforming, which cut synthesis-gas-compression costs and increased heat recovery. The first pressure reformer[3] was built in 1953. In addition, the higher pressures allowed improvements in the efficiency of synthesis-gas-purification systems. High-pressure steam-reforming technology also has been extended to cover heavier hydrocarbon gases, including propane, butane, reformer gases, and streams containing a high amount of olefins. In 1962, a process[4] for reforming straight-run liquid distillates (naphthas) was commercialized. This process is based on the use of an alkali oxide-promoted nickel catalyst[5] which permits reforming of desulfurized naphthas at low ($\sim$3.5:1) steam-to-carbon ratios, without significant carbon deposition problems.

Noncatalytic partial oxidation processes designed to produce H_2 from a wide range of hydrocarbon liquids, including heavy fuel oils, crudes, naphthas, coal tar, and pulverized bituminous coal, were commercialized in 1954[6] and 1956.[7] In both these processes, the hydrocarbon feed is oxidized and reformed in a refractory-lined pressure vessel. The required oxygen usually is supplied by an air separation plant from which nitrogen also is used as feed for the synthesis gas. The main differences between the two processes are in the reactor design, feeding method, burner design, and carbon and heat recovery. The partial oxidation processes and the steam-naphtha reforming process are favored in areas with short supplies of natural gas.

The source of nitrogen for the synthesis gas has always been air, either supplied directly from a liquid-air separation plant or by burning a small amount of the hydrogen with air in the H_2 gas. The need for air separation plants has been eliminated in modern ammonia plants by use of secondary reforming, where residual methane from the primary re-

---

[2] Originally developed by Standard Oil Company of New Jersey.

[3] Built by M. W. Kellogg for Shell Chemical Corp. (Ventura, California).

[4] M. W. Kellogg and Imperial Chemical Industries.

[5] Developed by M. W. Kellogg.

[6] Texaco partial oxidation process.

[7] Shell gasification process.

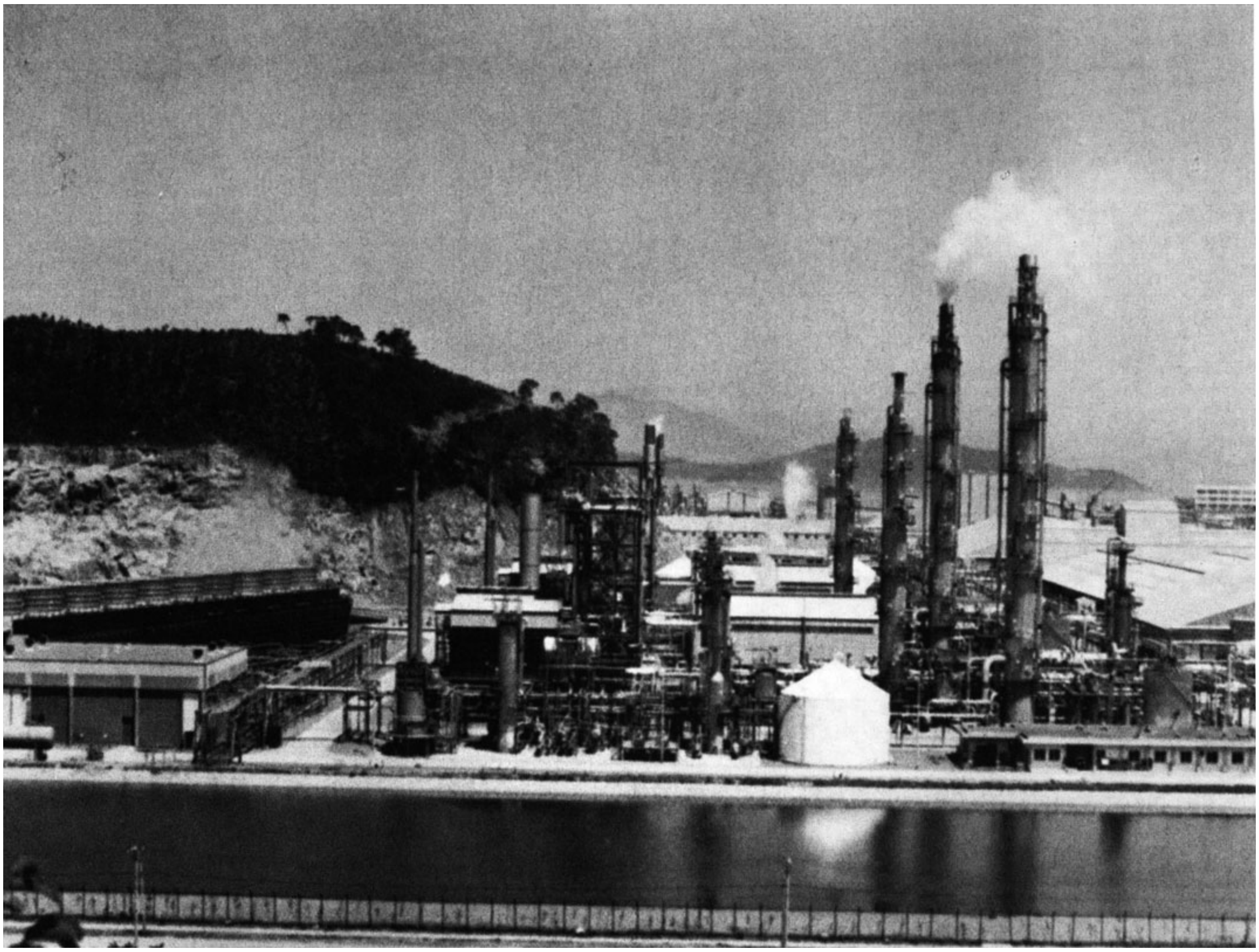


Fig. 2. Two 1000 short tons/day (900 metric tons/day) Kellogg-designed modern ammonia plants. (*M. W. Kellogg.*)

former is adiabatically reformed with sufficient air to produce a 3:1 mole ratio hydrogen-nitrogen synthesis gas.

Most ammonia plants built since the early 1960s are in the 600–1500 short tons/day (540–1350 metric tons/day) range and are based on new integrated designs that have cut the cost of ammonia manufacture in half. The plants of the early 1960s, in fact, have reached the best combination in terms of plant overall efficiency and cost by combining all the separate units (e.g., synthesis-gas preparation, purification, and ammonia synthesis) in one single train. High-pressure reforming has reduced the synthesis-gas compression load and front end plant equipment size. This compactness in design has also led to increased plant size at reduced investment and operating costs.

Use of Multistage Centrifugal Compressors. One of the major factors contributing to the improved economics of ammonia plants is the application of multistage centrifugal compressors, which have replaced the reciprocating compressors traditionally used in the synthesis feed and recycle service. A single centrifugal compressor can do the job of several banks of reciprocating compressors, thus reducing equipment cost, floor space, supporting foundations, and maintenance.

The use of multistage centrifugal compressors was made possible by redesigning the synthesis loop to operate at low pressures (150-240 atmospheres) and by increasing plant capacity to above the compressor's minimum-flow restriction in order to obtain a reasonable compressor efficiency. (Most synthesis loops using reciprocating compressors had been operating at intermediate pressures of 300–350 atmospheres.) Centrifugal compressors capable of developing pressures up to 340 atmospheres already are being offered and used in some large-capacity (1000 short tons/day; 900 metric tons/day) plants, where the increasing compressor horsepower is partially offset by reduction of the refrigeration horsepower requirement.

An operating ammonia plant using the aforementioned improvements is shown schematically in Fig. 1. This plant[8] has a capacity of 1000 short tons/day (900 metric tons/day) and uses natural gas as feedstock. The plant can be divided into the following integrated-process sections: (a) synthesis-gas preparation; (b) synthesis-gas purification; and (c) compression and ammonia synthesis. A typical (Kellogg designed) ammonia plant is shown in Fig. 2.

Synthesis Gas Preparation. The desulfurized natural gas mixed with steam is fed to the primary reformer, where it is reacted with steam in nickel-catalyst-filled tubes to produce a major percentage of the hydrogen required. The principal reactions taking place are[9]

$$CH_4 + H_2O \rightleftharpoons CO + 3H_2 \tag{1}$$

$$\Delta H_{298} = 49.3 \text{ kcal/mole}$$

$$CO + H_2O \rightleftharpoons CO_2 + H_2 \tag{2}$$

$$\Delta H_{298} = -9.8 \text{ kcal/mole}$$

Reaction (1) is the principal reforming reaction, and reaction (2) is the water-gas shift reaction. The net reactions are highly endothermic. The partially reformed gas leaves the primary reformer containing approximately 10% methane, on a mole dry-gas basis, at 27–34 atmospheres and up to 816°C. The required heat of reaction is supplied by natural-gas-fired arch burners, which are designed to also burn purge and flash

---

[8] Designed by The M. W. Kellogg Company, Houston, Texas, for which Kellogg received the 1967 Kirkpatrick Chemical Engineering Achievement Award.

[9] Heats of reaction at 198°K (25°C), 1 atmosphere pressure, gaseous substances in ideal state.

gases from the synthesis section. Waste heat from the primary reformer flue gas is recovered by generating high-pressure superheated steam, which along with waste-heat process boilers and an appended auxiliary boiler assure a steam system that is always in balance, while providing high-pressure steam to compressor turbine drivers and low-pressure steam to pump drivers. Further waste heat is recovered by preheating the natural-gas-steam feed mixture, steam-air for secondary reforming, and fuel.

The primary reforming step is followed by conversion of the residual methane to hydrogen and carbon oxides over a bed of high-temperature chrome and nickel catalysts in the secondary reformer. The secondary reforming step not only achieves a great degree of overall reforming economically possible, but also reduces fuel-gas input and overall reforming costs by shifting part of the required hydrocarbon conversion from the high-cost primary reformer to the lower-cost secondary reformer. It also permits an increase in the residual methane level at the primary effluent, which results in lower operating temperatures, reduced steam requirements, and milder tube-metal conditions.

Process waste-heat boilers then cool the reformed gas to about 371°C while generating high-pressure steam. The cooled gas-stream mixture enters a two-stage shift converter. The purpose of shift conversion is to convert CO to CO_2 and produce an equivalent amount of H_2 by the reaction: $CO + H_2O \rightleftharpoons CO_2 + H_2$. Since the reaction rate in the shift converter is favored by high temperatures, but equilibrium is favored by low temperatures, two conversion stages, each with a different catalyst provide the optimum conditions for maximum CO shift. Gas from the shift converter is the raw synthesis gas, which, after purification, becomes the feed to the NH_3 synthesis section.

Purification of Synthesis Gas. This involves the removal of carbon oxides to prevent poisoning of the NH_3 catalyst. An absorption process is used to remove the bulk of the CO_2, followed by methanation of the residual carbon oxides in the methanator. Modern ammonia plants use a variety of CO_2-removal processes with effective absorbent solutions. The principal absorbent solutions currently in use are hot carbonates and ethanolamines. Other solutions used include methanol, acetone, liquid nitrogen, glycols, and other organic solvents.

The partially purified synthesis gas leaves the CO_2 absorber containing approximately 0.1% CO_2 and 0.5% CO. This gas is preheated at the methanator inlet by heat exchange with the synthesis-gas compressor interstage cooler and the primary-shift converter effluent and reacted over a nickel oxide catalyst bed in the methanator. The methanation reactions are highly exothermic and are equilibrium favored by low temperatures and high pressures.

$$CO + 3H_2 \rightleftharpoons CH_4 + H_2O$$

$$CO_2 + 4H_2 \rightleftharpoons CH_4 + 2H_2O$$

The methanator effluent is cooled by heat exchange with boiler feedwater and cooling water. The synthesis gas leaves the methanator containing less than 10 parts per million (ppm) of carbon oxides.

Compression and Synthesis. The purified synthesis gas, containing H_2 and N_2 in a 3:1 mole ratio and with an inert gas (methane and argon) content of about 1.3 mole percent, is delivered to the suction of the synthesis-gas compressor. Anhydrous ammonia is catalytically synthesized in the converter. The effluent from the converter, after taking off a small purge stream, is recycled for eventual conversion to ammonia. Reaction takes place at approximately 370–482°C. Ammonia liquid, separated from the loop in the separator and from the purge, contains dissolved synthesis gas, which is released when the combined stream is flashed into the letdown drum. The flashed gas is then separated in the letdown drum and combined with the vapors from the purge separator to form a stream of purge fuel gases. Liquid ammonia in the letdown drum still contains some dissolved gases which must be disengaged. This liquid ammonia is let down to the refrigeration system where the dissolved gases are flashed and released to fuel. A centrifugal compressor is used to provide the refrigeration for the ammonia condensing. Ammonia product is withdrawn from the ammonia refrigeration system.

Modern Ammonia Plant. Since the development of the single train ammonia plant in the 1960s, many improvements have been made to reduce energy consumption. The plant of the 1980s has achieved striking results of reducing energy consumption by 20 to 30%—to less than 25 MMBTU(LHV)/ST (million Btus of low heating value fuel per short ton) of ammonia. This achievement represents a constant effort of development in seeking out more energy efficient design. Those developments have been centered around the 1960s basic process scheme with modifications to improve efficiency. Therefore, the basic process steps have not changed in any major way. The modern ammonia plants (1980s) have incorporated energy-saving features, including: (1) more efficient furnace design to reduce fuel consumption, (2) more efficient drivers, compressors, and reduced power consumption, (3) low energy consumption in carbon dioxide removal system, (4) more efficient waste heat recovery and utilization, and (5) more efficient synthesis loop design, such as make-up gas drying, purge gas hydrogen recovery, and intercooled ammonia converter. The trend toward greater energy conservation is expected to continue.

Future Considerations in Ammonia Production. In addition to continued emphasis on energy efficiency, alternate feedstocks will continue to be a primary area of ammonia technology.

In the past, coal or heavy hydrocarbon feedstock ammonia plants were not economically competitive with plants where the feedstocks were light hydrocarbons (natural gas to naphtha). Because of changing economics, however, plants that can handle heavy hydrocarbon feedstock are now attracting increasing attention. In addition, the continuous development and improvement of partial oxidation processes at higher pressure have allowed reductions in equipment size and cost. Therefore, the alternate feedback ammonia plants based on a partial oxidation process may become economically competitive in the near future.

The ultimate goal of any ammonia process will be the direct fixation of nitrogen by reaction of water with air, $1.5H_2O + 0.5N_2 \rightleftharpoons NH_3 + 0.750_2$. The theoretical energy requirement of the reaction is about 18 MMBTU/ST of ammonia. This feed energy requirement is the same as a natural gas feed ammonia plant. The major difference is that there is no short supply of water, air, and solar energy. However, the technology required for this route is not expected to be available any time soon. It is believed that, in the near future, ammonia plant designs will be based essentially on the present-day process with modifications to reduce energy consumption.

AMMONOLYSIS. See **Amination; Organic Chemistry.**

AMMONIUM CHLORIDE. NH_4Cl, formula weight 53.50, white crystalline solid, decomposes at 350°, sublimes at 520°C under controlled conditions, sp gr 1.52. Also known as *sal ammoniac*, the compound is soluble in H_2O and in aqueous solutions of NH_3; slightly soluble in methyl alcohol. Ammonium chloride is a high-tonnage chemical, finding uses as an ingredient of dry cell batteries, as a soldering flux, as a processing ingredient in textile printing and hide tanning, and as a starting material for the manufacture of other ammonium chemicals. The compound can be produced by neutralizing HCl with NH_3 gas or with liquid NH_4OH, evaporating the excess H_2O, followed by drying, crystallizing, and screening operations. The product also can be formed in the gaseous phase by reacting hydrogen chloride gas with NH_3. Ammonium chloride generally is not attractive as a source of nitrogen for fertilizers because of the build-up and damaging effects of chloride residuals in the soil. See also **Nitrogen.**

AMMONIUM COMPOUNDS. Several of the principal ammonium compounds are described in separate entries in this volume. See also **Ammonium Chloride; Ammonium Nitrate; Ammonium Phosphates;** and **Ammonium Sulfate.** The important aspects of several other ammonium compounds are summarized below.

Acetate: Ammonium acetate $NH_4C_2H_3O_2$, white solid, soluble, formed by reaction of ammonia or NH_4OH and acetic acid, reacts upon heating to yield acetamide.

Alum: Ammonium alums are those alums, such as aluminum ammonium sulfate $Al_2(NH_4)_2(SO_4)_4{\cdot}24H_2O$, ferric ammonium sulfate $Fe_2(NH_4)_2(SO_4)_4{\cdot}24H_2O$, chromium ammonium sulfate $Cr_2(NH_4)_2(SO_4)_4{\cdot}24H_2O$ where ammonium sulfate is crystallized with the heavier metal sulfate.

Benzoate: Ammonium benzoate $NH_4C_7H_5O_2$, white solid, soluble, formed by reaction of NH_4OH and benzoic acid. Used (1) as a food preservative, (2) in medicine.

Borate: Ammonium borate, ammonium tetraborate

$$(NH_4)_2B_4O_7 \cdot 4H_2O,$$

white solid, soluble, formed by reaction of NH_4OH and boric acid. Used (1) in fireproofing fabrics, (2) in medicine.

Bromide: Ammonium bromide NH_4Br, white solid, soluble, sublimes at 542°C, formed by reaction of NH_4OH and hydrobromic acid. Used in photography.

Carbonates: Ammonium carbonate, $(NH_4)_2CO_3$, volatile, white solid, soluble, formed by reaction of NH_4OH and CO_2 by crystallization from dilute alcohol, loses NH_3, CO_2, and H_2O at ordinary temperatures, rapidly at 58°C; ammonium hydrogen carbonate, ammonium bicarbonate, ammonium acid carbonate NH_4HCO_3, white solid, soluble, formed by reaction of NH_4OH and excess CO_2. This salt is the important reactant in the ammonia soda process for converting sodium chloride in solution into sodium hydrogen carbonate solid.

Chloroplatinate: Ammonium chloroplatinate $(NH_4)_2PtCl_6$, yellow solid, insoluble, formed by reaction of soluble ammonium salt solutions and chloroplatinic acid. Used in the quantitative determination of ammonium.

Cobaltinitrite: Diammonium sodium cobaltinitrite,

$$(NH_4)_2NaCo(NO_2)_6 \cdot H_2O$$

golden yellow precipitate, formed by reaction of sodium cobaltinitrite solution in acetic acid with soluble ammonium salt solution. Used in the detection of ammonium.

Cyanate: Ammonium cyanate NH_4CNO, white solid, soluble, formed by fractional crystallization of potassium cyanate and ammonium sulfate (ammonium cyanate is soluble in alcohol), when heated changes into urea.

Dichromate: Ammonium dichromate $(NH_4)_2Cr_2O_7$, red solid, soluble, upon heating evolves nitrogen gas and leaves a green insoluble residue of chromic oxide.

Fluoride: Ammonium fluoride NH_4F, white solid, soluble, formed by reaction of NH_4OH and hydrofluoric acid, and then evaporating. Used (1) as an antiseptic in brewing, (2) in etching glass; ammonium hydrogen fluoride, ammonium bifluoride, ammonium acid fluoride NH_4F_2, white solid, soluble.

Iodide: Ammonium iodide NH_4I, white solid, soluble, formed by reaction of NH_4OH and hydriodic acid, and then evaporating. Used (1) in photography, (2) in medicine.

Linoleate: Ammonium linoleate $NH_4C_{18}H_{31}O_2$. Used (1) as an emulsifying agent, (2) as a detergent.

Nitrite: Ammonium nitrite NH_4NO_2 when ammonium sulfate or chloride and sodium or potassium nitrite are heated, the mixture behaves like ammonium nitrite in yielding nitrogen gas.

Oxalate: Ammonium oxalate $(NH_4)_2C_2O_4$, white solid, soluble, formed by reaction of NH_4OH and oxalic acid, and then evaporating. Used as a source of oxalate; ammonium binoxalate $NH_4HC_2O_4 \cdot H_2O$, white solid, soluble.

Perchlorate: Ammonium perchlorate NH_4ClO_4, white solid, soluble, formed by reaction of NH_4OH and perchloric acid, and then evaporating. Used in explosives and pyrotechnics.

Periodate: Ammonium periodate NH_4IO_4, white solid, moderately soluble.

Persulfate: Ammonium persulfate $(NH_4)_2S_2O_8$, white solid, soluble, formed by electrolysis of ammonium sulfate under proper conditions. Used (1) as a bleaching and oxidizing agent, (2) in electroplating, (3) in photography.

Phosphomolybdate: Ammonium phosphomolybdate

$$(NH_4)_3PO_4 \cdot 12MoO_3$$

(or similar composition), yellow precipitate, soluble in alkalis, formed by excess ammonium molybdate and HNO_3 with soluble phosphate solution. Used as an important test for phosphate (similar product and reaction when arsenate replaces phosphate).

Salicylate: Ammonium salicylate $NH_4C_7H_5O_3$, white solid, soluble, formed by reaction of NH_4OH and salicylic acid, and then evaporating. Used in medicine.

Sulfide: Ammonium sulfide $(NH_4)_2S$, colorless to yellowish solution, formed by saturation with hydrogen sulfide of one-half of a solution of NH_4OH, and then mixing with the other half of the NH_4OH. Dissolves sulfur to form ammonium polysulfide, yellow solution. Used as a reagent in analytical chemistry; ammonium hydrogen sulfide, ammonium bisulfide, ammonium acid sulfide NH_4HS, colorless to yellowish solution, formed by saturation with H_2S of a solution of NH_4OH.

Tartrate: Ammonium tartrate $(NH_4)_2C_4H_4O_6$, white solid, moderately soluble, formed by reaction of NH_4OH and tartaric acid, and then evaporating. Used in the textile industry; ammonium hydrogen tartrate, ammonium bitartrate, ammonium acid tartrate $NH_4HC_4H_4O_6$, white solid, slightly soluble, formation sometimes used in detection of ammonium or tartrate.

Thiocyanate: Ammonium thiocyanate, ammonium sulfocyanide, ammonium rhodanate NH_4CNS, white solid, soluble, absorbs much heat on dissolving with consequent marked lowering of temperature, mp 150°C, formed by boiling ammonium cyanate solution with sulfur, and then evaporating. Used (1) as a reagent for ferric, (2) in making cooling solutions, (3) to make thiourea.

Ammonium compounds liberate NH_3 gas when warmed with NaOH solution.

AMMONIUM HYDROXIDE. NH_4OH, formula weight 35.05, exists only in the form of an aqueous solution. The compound is prepared by dissolving NH_3 in H_2O and usually is referred to in industrial trade as aqua ammonia. For industrial procurements, the concentration of NH_3 in solution is normally specified in terms of the specific gravity (degrees Baumé, °Be). Common concentrations are 20° Be and 26° Be. The former is equivalent to a sp gr of 0.933, or a concentration of about 17.8% NH_3 in solution; the latter is equivalent to a sp gr of 0.897, or a concentration of about 29.4% NH_3. These figures apply at a temperature of 60°F (15.6°C). Reagent grade NH_4OH usually contains approximately 58% NH_4OH (from 28 to 30% NH_3 in solution).

Ammonium hydroxide is one of the most useful forms in which to react NH_3 (becoming the NH_4^+ radical in solution) with other materials for the creation of ammonium salts and other ammonium and nitrogen-bearing chemicals. Ammonium hydroxide is a direct ingredient of many products, including saponifiers for oils and fats, deodorants, etching compounds, and cleaning and bleaching compounds. Because aqua ammonia is reasonably inexpensive and a strongly alkaline substance, it finds wide application as a neutralizing agent. See also **Nitrogen.**

AMMONIUM NITRATE. NH_4NO_3, formula weight 80.05, colorless crystalline solid, occurs in two forms:

α-NH_4NO_3, tetragonal crystals, stable between -16°C and 32°C, sp gr 1.66.

β-NH_4NO_3, rhombic or monoclinic crystals, stable between 32°C and 84°C, sp gr 1.725.

The melting point generally ascribed to the alpha form is 169.6°C, with decomposition occurring above 210°C. Upon heating, ammonium nitrate yields nitrous oxide (N_2O) gas and can be used as an industrial source of that gas. Ammonium nitrate is soluble in H_2O, slightly soluble in ethyl alcohol, moderately soluble in methyl alcohol, and soluble in acetic acid solutions containing NH_3.

As shown in the accompanying figure, in making ammonium nitrate on a large scale, NH_3, vaporized by waste steam from neutralizer, is sparged along with HNO_3 into the neutralizer. A ratio controller automatically maintains the proper proportions of NH_3 and acid. The heat of neutralization evaporates a part of the H_2O and gives a solution of 83% NH_4NO_3. Final evaporation to above 99% for agricultural prills or to approximately 96% for industrial prills is accomplished in a falling-film evaporator located at the top of the prilling tower. The resultant melt flows through spray nozzles and downward through the tower. Air is drawn upward by fans at the top of the tower. The melt is cooled sufficiently to solidify, forming round pellets or prills of the desired range of sizes. The prills are removed from the bottom of the tower and fed to a rotary cooler. Where industrial-type prills are produced, a pre-

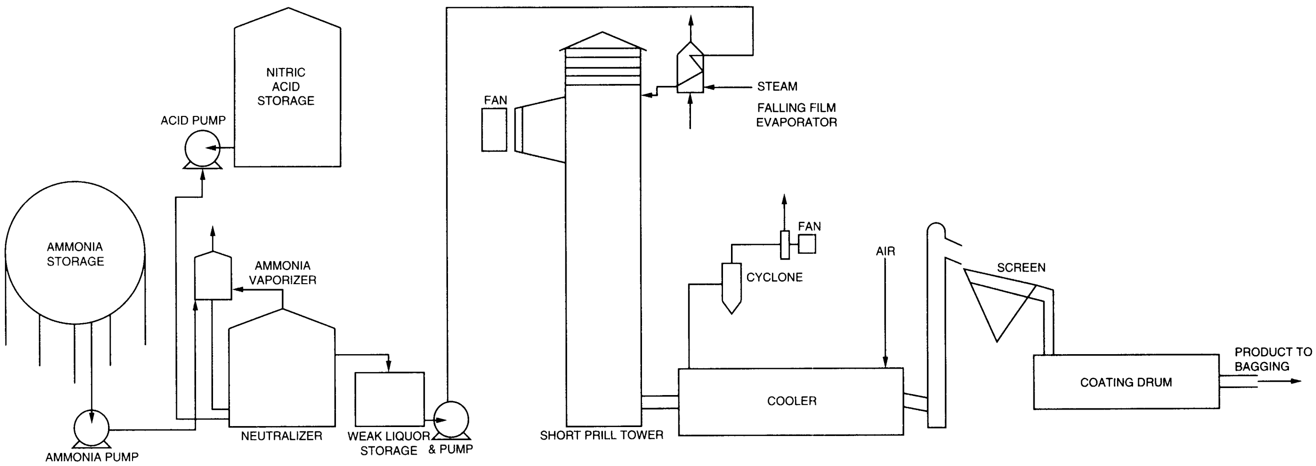


Fig. 1. Process for making ammonium nitrate on a large scale.

drier and drier precede the cooler. Fines from the rotary drums are collected in wet cyclones. This solution eventually is returned to the neutralizer. After cooling, the prills are screened to size and the over-and undersize particles are sent to a sump and returned to the neutralizer. Intermediate or product-size prills are dusted with a coating material, usually diatomaceous earth, in a rotary coating drum and sent to the bagging operation. The process can be adapted to other types of materials and mixtures of ammonium nitrate and other fertilizer materials. Mixtures include the incorporation of limestone and ammonium phosphates.

Ammonium nitrate is a very high tonnage industrial chemical, finding major applications in explosives and fertilizers, and additional uses in pyrotechnics, freezing mixtures (for obtaining low temperatures), as a slow-burning propellant for missiles (when formulated with other materials, including burning-rate catalysts), as an ingredient in rust inhibitors (especially for vapor-phase corrosion), and as a component of insecticides.

Amatol, an explosive developed by the British, is a mixture of ammonium nitrate and TNT. A special explosive for tree-trunk blasting consists of ammonium nitrate coated with TNT. In strip mining, an explosive consisting of ammonium nitrate and carbon black is used. The explosive ANFO is a mixture of ammonium nitrate and fuel oil. ANFO accounts for about 50% of the commercial explosives used in the United States. Slurry explosives consist of oxidizers (NH_4NO_3 and $NaNO_3$), fuels (coals, oils, aluminum, other carbonaceous materials), sensitizers (TNT, nitrostarch, and smokeless powder), and water mixed with a gelling agent to form a thick, viscous explosive with excellent water-resistant properties. Slurry explosives may be manufactured as cartridged units, or mixed on-site. Although Nobel introduced NH_4NO_3 into his dynamite formulations as early as 1875, the tremendous explosive power of the compound was not realized until the tragic Texas City, Texas disaster of 1947 when a shipload of NH_4NO_3 blew up while in harbor. See also **Explosive.**

As a fertilizer, NH_4NO_3 contains 35% nitrogen. Because of the explosive nature of the compound, precautions in handling are required. This danger can be minimized by introducing calcium carbonate into the mixture, reducing the effective nitrogen content of the product to 26%. In as much as NH_4NO_3 is hightly hygroscopic, clay coatings and moisture-proof bags are means used to preclude spoilage in storage and transportation. See also **Fertilizer;** and **Nitrogen.**

AMMONIUM PHOSPHATES. There are two ammonium phosphates, both produced on a very high-tonnage scale.

Monoammonium phosphate, $NH_4H_2PO_4$, white crystals, sp gr 1.803
Formula weight 115.04, $N = 12.17\%$, $P_2O_5 = 61.70\%$
Diammonium phosphate, $(NH_4)_2HPO_4$, white crystals, sp gr 1.619
Formula weight 132.07, $N = 21.22\%$, $P_2O_5 = 53.74\%$

Both compounds are soluble in H_2O; insoluble in alcohol or ether. A third compound, triammonium phosphate $(NH_4)_3PO_4$ does not exist under normal conditions because, upon formation, it immediately decomposes, losing NH_3 and reverting to one of the less alkaline forms.

Large quantities of the ammonium phosphates are used as fertilizers and in fertilizer formulations. The compounds furnish both nitrogen and phosphorus essential to plant growth. The compounds also are used as fire retardants in wood building materials, paper and fabric products, and in matches to prevent afterglow. Solutions of the ammonium phosphates sometimes are air dropped to retard forest fires, serving the double purpose of fire fighting and fertilizing the soil to accelerate new plant growth. The compounds are used in baking powder formulations, as nutrients in the production of yeast, as nutritional supplements in animal feeds, for controlling the acidity of dye baths, and as a source of phosphorus in certain kinds of ceramics.

Ammonium phosphates usually are manufactured by neutralizing phosphoric acid with NH_3. Control of the pH (acidity/alkalinity) determines which of the ammonium phosphates will be produced. Pure grades can be easily made by crystallization of solutions obtained from furnacegrade phosphoric acid. Fertilizer grades, made from wet-process phosphoric acid, do not crystallize well and usually are prepared by a granulation technique. First, a highly concentrated solution or slurry is obtained by neutralization. Then the slurry is mixed with from 6 × to 10 × its weight of previously dried material, after which the mixture is dried in a rotary drier. The dry material is then screened to separate the desired product size. Oversize particles are crushed and mixed with fines from the screen operation and then returned to the granulation step where they act as nuclei for the production of further particles. Other ingredients often are added during the granulation of fertilizer grades. The ratio of nitrogen to phosphorus can be altered by the inclusion of ammonium nitrate, ammonium sulfate, or urea. Potassium salts sometimes are added to provide a 3-component fertilizer (N, P, K). A typical fertilizer grade diammonium phosphate will contain 18% N and 46% P_2O_5 (weight). See also **Fertilizer;** and **Nitrogen.**

There has been a trend toward the production of ammonium phosphates in powder form. Concentrated phosphoric acid is neutralized under pressure, and the heat of neutralization is used to remove the water in a spray tower. The powdered product then is collected at the bottom of the tower. Ammonium nitrate/ammonium phosphate combination products can be obtained either by neutralizing mixed nitric acid and phosphoric acid, or by the addition of ammonium phosphate to an ammonium nitrate melt.

AMMONIUM SULFATE. $(NH_4)_2SO_4$, formula weight 132.14, colorless crystalline solid, decomposes above 513°C, sp gr 1.769. The compound is soluble in H_2O and insoluble in alcohol. Ammonium sulfate is a high-tonnage industrial chemical, but frequently may be considered a by-product as well as intended end-product of manufacture. Large quantities of ammonium sulfate result from a variety of industrial neutralization operations required for alleviation of stream pollution by free H_2SO_4. The ammonium sulfate so produced is not always recovered and marketed. A significant commercial source of $(NH_4)_2SO_4$ is its creation as a by-product in the manufacture of caprolactam, which yields several tons of the compound per ton of caprolactam made. See also **Caprolactam**. Ammonium sulfate also is a by-product of coke oven operations where the excess NH_3 formed is neutralized with H_2SO_4 to form $(NH_4)_2SO_4$. However, as a major fertilizer and ingredient of fertilizer formulations, additional production is required, largely depending upon the proximity of consumers to by-product $(NH_4)_2SO_4$ sources. In the Meresburg reaction, natural or by-product gypsum is reacted with ammonium carbonate:

$$CaSO_4{\cdot}2H_2O + (NH_4)_2CO_3 \rightarrow CaCO_3 + (NH_4)_2SO_4 + 2H_2O.$$

The product is stable, free-flowing crystals. As a fertilizer, $(NH_4)_2SO_4$ has the advantage of adding sulfur to the soil as well as nitrogen. By weight, the compound contains 21% N and 24% S. Ammonium sulfate also is used in electric dry cell batteries, as a soldering liquid, as a fire retardant for fabrics and other products, and as a source of certain ammonium chemicals. See also **Fertilizer;** and **Nitrogen.**

AMNESIA. A partial or total loss of memory of a temporary or permanent nature. The condition may result from a brain injury (See **Brain (Injury)**); or it may be symptomatic of a basic disorder of the mind. The condition is frequently associated with a dissociative reaction, but may also be evidenced in other psychoses resulting from stress, such as anxiety, phobic, obsessive-compulsive, and conversion reactions. *Retrograde amnesia* is an impaired ability to recall past events; *anterograde amnesia* is an impaired ability to learn new information; *confabulation*, sometimes associated with amnesia, is the fabrication of recent events.

In retrograde amnesia, anxiety over an event becomes so great that the individual is forced to forget it. In forgetting anxiety, the patient also forgets a multitude of necessary associations, including personal identity. Despite these shortcomings, the individual is often well oriented as to present time and place. The person simply cannot recall anything about the past. The patient's behavior appears so normal that the individual moves about freely without attracting undue notice. In some instances, the individual may wander restlessly from place to place, covering extensive regions in such travels.

Recovery of memory is sometimes spontaneous. Frequently, psychiatric help is required. This assistance may have to extend over an appreciable period. Upon regaining memory, the amnesic patient does not recall events which occurred during the period of amnesia. However, it is frequently possible to bring forth recent past events in considerable detail through hypnosis. This indicates that the loss of consciousness in dissociation is different from that of the delirious states. Patients who recover from delirium cannot recall experiences even when hypnotized.

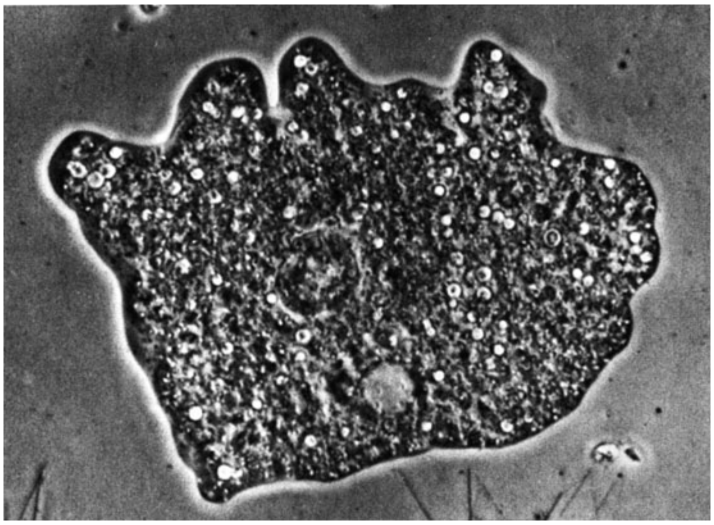
In a dissociative reaction, there is an unconscious flight from situations of intolerable emotional stress. The reaction takes numerous forms, including some kinds of amnesia, sleep-walking, automatic writing, and the extremely rare "dual personality," in which two mental selves exist within the same body at the same time.

Amnesia may also occur as a consequence of Wernicke's encephalopathy. This condition is found in some chronic alcoholics whose diet is inadequate (frequently characterized by lack of thiamine). Neuronal degeneration results in mental confusion, disorders of eye movement (gaze), and ataxia. When the condition is left untreated, permanent neurological damage occurs, resulting in a severe impairment of memory known as Korsakoff's psychosis (or amnesic-confabulatory psychosis).

AMNIOCENTESIS. Extraction and analysis of some of the amniotic fluid from the sac surrounding the fetus. See **Embryo.**

AMNION. An accessory embryonic membrane common to reptiles, birds and mammals and a superficially similar structure found in some insects.

AMOEBA. A genus of one-celled animals in which the body consists of a naked mass of protoplasm and the organs of locomotion are temporary blunt protuberances of cytoplasm known as pseudopodia. The large fresh water form, *Amoeba proteus*, is a typical species. When seen under the microscope, this protozoan appears as a naked bit of protoplasm surrounded only by its thin plasma membrane. When in an active state, it is constantly changing its shape as it moves about. See also **Asexual Reproduction; Cell (Biology).**



Amoeba. (*A. M. Winchester.*)

AMOEBOID MOVEMENT. Movement of cells by means of pseudopodia, as in *Amoeba*. The white blood cells, leucocytes, in the blood of higher animals are good examples of cells that move in this manner.

AMOEBULAE. Spores of amoeboid form which are produced by some one-celled animals.

AMOR ASTEROIDS. See **Asteroid.**

AMORPHOUS. As opposed to a crystalline substance which exhibits an orderly structure, the behavior of an amorphous substance is similar to a very viscous, inelastic liquid. Examples of amorphous substances include amber, glass, and pitch. An amorphous material may be regarded as a liquid of great viscosity and high rigidity, with physical properties the same an all directions (may be different for crystalline materials in different directions). Usually, upon heating, an amorphous solid gradually softens and acquires the characteristics of a liquid, but without a definite point of transition from solid to liquid state. In geology, an amorphous mineral lacks a crystalline structure, or has an internal arrangement so irregular that there is no characteristic external form. This does not preclude, however, the existence of any degree of order. The term amorphous is used in connection with amorphous graphite and amorphous peat, among other naturally occurring substances.

AMOSITE. Amosite is a long-fiber gray or greenish asbestiform mineral related to the cummingtonite-grunerite series, and is of economic importance. It occurs within both regional and contact metamorphic rocks in the Republic of South Africa. The name *amosite* is a product of the initial letters of its occurrence at the Asbestos Mines of South Africa. See also **Asbestos.**

AMPERE. See **Units and Standards.**

AMPERE'S LAW. This law of magnetostatics has been stated in a number of forms, one in terms of the magnetic field intensity produced by a current flowing in a thin conductor and another in terms of a line

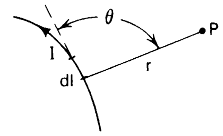


Fig. 1. Ampere's law in terms of magnetic field intensity.

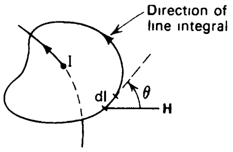


Fig. 2. Ampere's law in terms of a line integral about a closed path.

integral about a closed path. The first form, which lacks exactness, is (refer to Fig. 1) that the magnitude of the magnetic field intensity dH at point P produced by a current of I amperes flowing through an element dl of a thin conducting wire at a distance r is

$$dH = \frac{Idl \cos \theta}{4\pi r^2} \text{ amperes/meter}$$

The direction of dH is given by the right-hand screw rule. The total magnetic field intensity at P is the sum of the vector fields produced by all the conductor elements. The second form (refer to Fig. 2) is that the line integral of the magnetic field intensity around a closed path is equal to the total current which flows through any surface bounded by the closed path or $\oint H \cos \theta \, dl = I$. All units are in the meter-kilogram-second system.

AMPERE'S RULE. The magnetic flux generated by a current in a wire encircles the current in the counterclockwise direction, if the current is approaching the observer.

AMPERE'S THEOREM. The magnetic field due to an electric current flowing in any circuit is equivalent at external points to that due to a simple magnetic shell, the bounding edge of which coincides with the conductor, and the strength of which is equal to the strength of the current.

AMPEROMETER. An instrument for the chemical analysis of electro-reducible or oxidizable ions, molecules, or dissolved gases in solution. Included are the majority of metal ions and many organic substances that contain oxidizable or reducible groups. The range of concentration measurements of which the instrument is capable is from 0.01 to 1,000 ppm. This instrumental method is used for determining free available or total available chlorine, particularly in connection with water-chlorination control. In this application, the range of chlorine concentration is from 0 to 50 ppm. Free iodine also may be determined with an amperometer. The identification of unknown substances is by inference from noting the fixed potential between polarized microelectrode and reference electrode that causes oxidation or reduction of a given composition sought. Where identity is firmly established, the concentration also may be determined because the concentration is proportional to the diffusion-limited current that flows in the electrode circuit. In this latter respect, amperometry is similar to polarography. See also **Analysis (Chemical).**

AMPHETAMINE. Also called methylphenethylamine; 1-phenyl-2-aminopropane; Benzedrine; formula $C_6H_5CH_2CH(NH_2)CH_3$; *amphetamine* is a colorless, volatile liquid with a characteristic strong odor and slightly burning taste. Boils and commences decomposition at 200–203°C. Low flash point, 26.7°C. Soluble in alcohol and ether; slightly soluble in water. Amphetamine is the basis of a group of hallucinogenic, habit-forming drugs which affect the central nervous system. The drug also finds medical application, notably in appetite suppressants. It should be emphasized that administration of amphetamines for prolonged periods in connection with weight-reduction programs may lead to drug dependence. Particular attention must be paid by professionals to the possibility of persons obtaining amphetamines for nontherapeutic use or distribution to others.

AMPHIBIA. The frogs, toads, newts, salamanders and related forms. A class of the phylum *Chordata*. Since these animals live only in moist places their distribution is restricted and they are among the less familiar vertebrates.

The amphibians are distinguished by: (1) moist skin; (2) the absence of scales and claws; (3) a metamorphosis, undergone by most species during development, from an aquatic, gill-breathing larva to a semi-terrestrial, air-breathing adult stage.

While some of the salamanders are permanently aquatic and some of the tree frogs permanently terrestrial, most members of the class live near the water or in moist places and undergo the metamorphosis mentioned above.

The following orders of amphibians are recognized:

Order *Gymnophiona (Apoda)*. Legless, worm-like animals, confined to the tropics of the Old and New Worlds.

Order *Urodela (Caudata)*. Elongate animals with long tails and weak, short legs. The salamanders, newts, efts, hellbender, and mud puppy.

Order *Anura (Salientia)*. Tailless species whose hind legs are the larger pair, more or less strongly developed for jumping. Most species have a larval stage known as the tadpole with a compact body and a long compressed tail but no legs until the onset of metamorphosis. The frogs and toads.

AMPHIBOLE. This is the name given to a closely related group of minerals all showing in common a prismatic cleavage of 54–56° as well as similar optical characteristics and chemical composition.

The amphiboles may be said to represent chemically a series of metasilicates corresponding to the general formula $RSiO_3$ where R may be calcium, magnesium, iron, aluminum, titanium, sodium, or potassium. The crystals of the amphibole family group fall within both the monoclinic and orthorhombic systems.

There is a clear parallelism between the amphiboles and the pyroxenes. There are two basic differences between the minerals of these two family groups; amphiboles with cleavage angles of 56° and 124°, with essential OH groups in their structure; pyroxenes with cleavage angles of 87° and 93°, and being anhydrous, with no OH content. Amphibole crystals are usually long and slender and tend to be simple while pyroxene crystals tend to be complex, short, and stout prisms.

Amphibole is common in both lavas and deep-seated rocks, though less so in the basic lavas than pyroxene. Many of the amphiboles may be developed as metamorphic minerals. The following members of the amphibole group are described under their own headings: actinolite, anthophyllite, cummingtonite, glaucophane, grünerite, hornblende, riebeckite and tremolite. Amphibole was so named by Haüy from the Greek word, meaning doubtful, because of the many varieties of this mineral. See also **Pyroxene.**

AMPHIBOLITE. The amphibolites form a large group of rather important rocks of metamorphic character. As the name implies they are made up very largely of minerals of the amphibole group. There may be also a variety of other minerals present, such as quartz, feldspar, biotite, muscovite, garnet, or chlorite in greater or lesser amounts.

Depending upon the particular amphibole present these rocks may be light to dark green or black, the amphibole usually being in long slender prisms or laths, often quite coarse, sometimes in acicular or fibrous forms.

Because the mineral constituents are arranged parallel to the schistosity, amphibolites may have a strongly developed cleavage.

The occurrence of amphibolites accompanying gneisses, schists, and other metamorphic rocks of probable sedimentary origin strongly suggests a similar derivation. Yet some amphibolites cut other metamorphic rocks in the manner of dikes or sills. It is very likely that they have been derived from both original igneous and sedimentary rocks. Large masses of amphibolite suggest gabbroic stocks. Well-known areas in which amphibolites are found are New England, New York State, Canada, Scotland, and the Alps.

AMPHIDROMIC POINT. A geographic position in the ocean where theoretically there is no tide range and from which cotidal lines radiate in various directions. The tide amplitude presumably increases with distance from this point. A synonym is *nodal point.*

AMPHINEURA. The chitons and allied forms, a group of the phylum *Mollusca*. The more familiar members are flattened marine animals of oval outline. They have a shell composed of a series of separate plates, sometimes concealed within the body. The foot makes up most of the ventral surface and the limited mantle extends down about it to form a shallow groove. Nerve cells are in many cases distributed through the nerve cords so that the nervous system contains no ganglia. See **Ganglion.**

The group has two classes:

Class *Aplacophora (Solenogastres)*. Worm-like animals, somewhat cylindrical and elongate. Shell lacking.

Class *Polyplacophora*. The chitons. Flattened and oval, with shell plates. They are widely distributed, chiefly in the shallow waters. Sometimes used as food.

AMPHIOXUS. Commonly used to designate any of the primitive chordates called lancelets but more accurately a genus of these animals. See **Cephalochordata.**

AMPHIPODA. An order of crustaceans including marine and freshwater species. The beach-fleas are among the few which have a common name.

AMPHIPROTIC. Capable of acting either as an acid or as a base, i.e., as a proton donor or acceptor, according to the nature of the environment. Thus, aluminum hydroxide dissolves in acids to form salts of aluminum, and it also dissolves in strong bases to form aluminates. Solvents like water which can act to give protons or accept them, are amphiprotic solvents. See **Acids and Bases;** and **Salt.**

AMPLIFICATION. A general transmission term used to denote an increase of signal magnitude.

AMPLIFIER. A device for increasing the strength of a signal without appreciably altering other signal characteristics, such as waveform. Amplifiers may be classified in several ways—by basic mode of operation, such as electronic (solid state or vacuum tube), magnetic, hydraulic, fluidic, and mechanical (level systems); by application, such as data and information systems (computing, electronic data processing, communications, instrumentation) and automatic control systems. In one device or subassembly, other functions may be combined with amplification. The nomenclature of amplifiers reflects these variations in terms of numerous special designations—some by mode of operation as a transistor amplifier, magnetic amplifier, fluidic amplifier, etc.; some by use and specific function as a carrier amplifier, sample-and-hold amplifier, servoamplifier, power amplifier, etc. Amplifiers also may be classified in terms of their frequency response, or in terms of the signal frequencies they are designed to transmit. This classification is essentially one which depends on the frequency characteristics of the load impedance or coupling network used between transistor amplifier stages. Thus, there are direct current (dc) amplifiers, audio frequency amplifiers, rf amplifiers, video frequency amplifiers, etc. Also, recognizing the mechanism of amplification as the control of an output current by an input voltage or current, an amplifier may be classified in terms of the fraction of the cycle of an assumed sinusoidally varying input signal for which the output current flows.

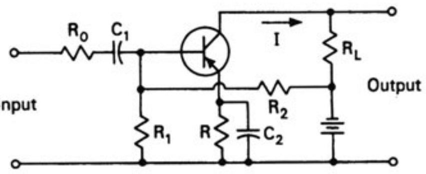


Fig. 1. Common emitter transistor amplifier connection.

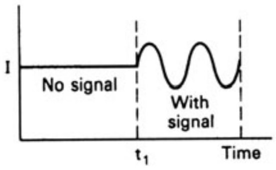


Fig. 2. Waveform in resistance-coupled amplifier.

A simple transistor amplifier is shown in Fig. 1. Capacitor C_1 is a blocking capacitor to prevent modification of the bias conditions due to a dc path through the generator, R_1 and R_2 in conjunction with the battery and the dc voltage developed across R produce the desired bias current in the base lead, and C_2 is a bypass capacitor to eliminate feedback due to signal current flowing through resistor R. The application of an input signal changes the current into the base-emitter junction with a consequent change in collector current I as shown in Fig. 2. The output voltage is developed across resistor R_L as a result of the flow of collector current through it. Resistor R_0 produces a load for the source of signal voltage which is far more constant than the base to emitter resistance of the transistor alone. The input signal current is made much less dependent on the device itself by this means. When the gain obtainable from a single amplifier is not sufficient for a given application, it is necessary to cascade two or more stages. This situation requires coupling networks between amplifier stages which assume a variety of forms depending on the frequency response characteristics desired for the amplifier. A simple and common method is known as resistance coupling. This is coupling in which resistors are used as the input and output impedances of the circuits being coupled. A coupling capacitor can be used between the resistors to transfer the signal from one stage to the next.

Amplifier Configurations. Several specific amplifier configurations are described briefly in the following listing, arranged alphabetically for convenience.

(*Balanced Amplifier*)—An amplifier circuit in which there are two identical signal branches connected so as to operate in phase opposition and with input and output connections each balanced to ground.

(*Booster Amplifier*)—An amplifier used in audio consoles between mixer controls and the master volume control to prevent deterioration of signal-to-noise ratio. It generally supplies sufficient gain to compensate for mixing-circuit losses.

(*Bootstrap Amplifier*)—A single-stage amplifier in which a change in input signal voltage changes the potential of the input source with respect to ground (or other reference voltage) by an amount equal to the output signal voltage. In a transistor amplifier, the input is applied between base and emitter with the output load connected between the emitter and the low potential side of the collector power supply.

(*Bridge Amplifier*)—Extensively used for instrumentation purposes, the commercial configuration generally is a direct-coupled amplifier, offering reasonably wide bandwidths up to 50 kHz at gains ranging from near unity to 1,000. See also separate entry on **Bridge Amplifier.**

(*Buffer Amplifier*)—This term is commonly applied to an amplifier stage whose main function is to isolate the oscillation of a transmitter from the main power amplifiers, but it is also applied to any amplifier which is inserted between two circuits or amplifier stages to reduce substantially the interaction of one on the other. The frequency of an oscillator depends, among other factors, upon the load which is applied to it. Since, for satisfactory communications, it is highly desirable that the oscillator frequency remain constant, buffer amplifiers are always used in broadcast transmitters and are usually used in others. Such an amplifier usually operates with the signal current to the input circuit of the buffer amplifier so low that the current taken from the exciting circuit is not sufficient to change its operation appreciably from the no-load condition. The output circuit of the buffer can then supply the normal load of a conventional amplifier without this load affecting the oscillator. Sometimes more than one buffer stage is used to make certain that no load is reflected back to the oscillator. See also **Buffer (Computer).**

(*Carrier Amplifier*)—A dc amplifier wherein the signal first is modulated, then demodulated during amplification. The bandwidth usually is from zero to some finite frequency. Electronic switches or electromechanical devices are used in most cases to effect the modulation. Thus, the "chopping" action accomplishes the equivalent of a square-wave modulation of the signal. See also separate entry on **Carrier Amplifier.**

(*Cascade Amplifier*)—A series of amplifiers with each output connected to the input of another amplifier (except the last).

(*Chopper Amplifier*)—In one type of chopper amplifier, the input signal is chopped (or modulated), amplified by an ac amplifier, demodulated, and filtered to provide a dc output signal. See foregoing mentioned description of carrier amplifier. In a second type, the error signal is chopped for the purpose of providing stabilization of gain and offset. The term, chopper-stabilized amplifier, may be a more apt designation. See also separate entry on **Chopper Amplifier.**

(*Common-Base Amplifier*)—A form of transistor amplifier in which the input signal is applied between emitter and base and the output signal is taken between collector and base. See Fig. 3. Characteristics associated with this connection are extremely low input impedance, a high output impedance, and a current amplification somewhat less than unity. The circuit is useful in providing impedance transformation from a low impedance source to a high impedance load.

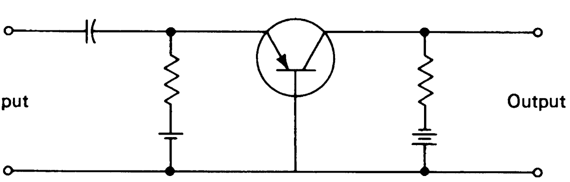


Fig. 3. Common-base transistor amplifier.

(*Common-Emitter Amplifier*)—A form of transistor amplifier in which the input signal is applied between base and emitter and the output signal is taken between collector and emitter. This configuration is capable of providing both current gain and voltage gain exceeding unity in contrast with the characteristics of the common-base and emitter-follower amplifiers. See Fig. 4.

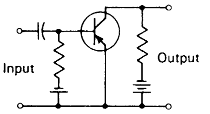


Fig. 4. Common-emitter amplifier.

(*Contact-Modulated Amplifier*)—Used for the amplification of dc and very low frequency signals. The signal source is modulated by a carrier-operated contact system (usually 60 to 400 Hz), the resulting modulated wave amplified in an ac amplifier to a suitable level, and subsequently demodulated, sometimes by the same contact system used to accomplish the original modulation.

(*Direct-Coupled Amplifier*)—An amplifier in which no coupling capacitors are employed as interstage coupling elements and which thus is capable of amplifying dc variations and ac signals of arbitrarily low frequency. A transistor direct-coupled amplifier using feedback is shown in Fig. 5. The need for batteries or Zener diodes as coupling elements has been avoided in this design by suitable choice of transistor operating points, values of collector and emitter resistors, and the feedback connections.

(*Distributed Amplifier*)—An amplifier consisting of components appropriately distributed along artificial transmission lines. This amplifier is capable of much greater bandwidths than a conventional amplifier, and the ordinary figure of merit or gain-bandwidth product does not apply.

(*Doherty Amplifier*)—A particular arrangement of a radio-frequency linear power amplifier wherein the amplifier is divided into two sections whose inputs and outputs are connected by quarter-wave (90°) networks, and whose operating parameters are so adjusted that, for all values of the input signal voltage up to one-half maximum amplitude, Section No. 2 is inoperative and Section No. 1 delivers all the power to the load, which presents an impedance at the output of Section No. 1 that is twice the optimum for maximum output. At one-half maximum input level, Section No. 1 is operating at peak efficiency, but is beginning to saturate. Above this level, Section No. 2 comes into operation, thereby decreasing the impedance presented to Section No. 1, which causes it to deliver additional power into the load until, at maximum signal input, both sections are operating at peak efficiency, and each section is delivering one-half the total output power to the load.

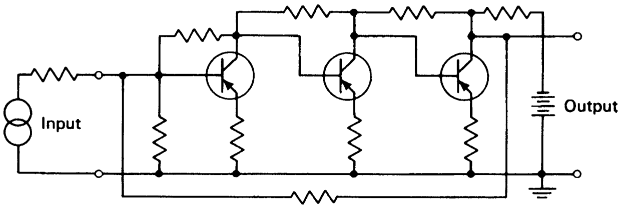


Fig. 5. Direct-coupled transistor amplifier.

(*Double-Stream Amplifier*)—A traveling-wave amplifier in which the amplification occurs as a result of the interaction of two electron beams having different average velocities. The amplification takes place in the beam itself and is a result of what might be called electromechanical interaction.

(*Feedback Amplifier*)—An amplifier in which feedback has been deliberately introduced to obtain certain performance characteristics. Commonly, the feedback connection is made from the output to the input of the amplifier and is employed to reduce variations in gain associated with changes in the characteristics of transistors. A representative feedback amplifier is shown in Fig. 6. In this case, the amplifier acts to maintain the output voltage constant for a constant peak amplitude input signal. The amplifier contains stabilizing networks (N_1) to prevent oscillation.

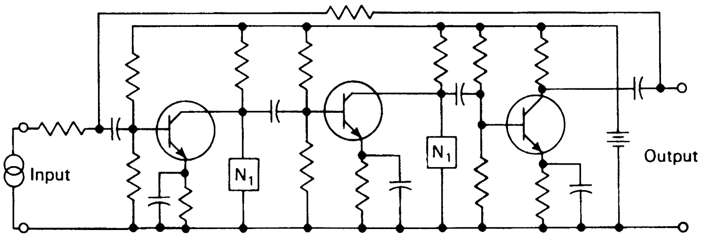


Fig. 6. Transistor feedback amplifier.

(*Floating Amplifier*)—Also known as an isolated amplifier, the design does not require that the input and output signals be referred to the same signal reference point (ground). Generally, differential-input amplifiers meet this definition. The term *floating*, however, normally excludes an amplifier where the input reference point and the output reference point are common, that is, either through a signal conductor or a power supply. Specifically, floating amplifier refers to an amplifier which includes a 4-terminal coupling device, such as a light-coupled signal-transmission element or a transformer. See also separate entry on **Floating Amplifier.**

(*Fluidic Amplifier*)—A fluid amplifier amplifies pressure, mass flow, and fluid power and, in principle, bears considerable resemblance to its electrical/electronic counterparts. Several types of fluidic amplifiers have been designed. See also separate entry on **Fluidics.**

(*Hydraulic Amplifier*)—A power amplifier employed in some servomechanisms and control systems in which power amplification is obtained by the control of the flow of a high-pressure liquid by a valve mechanism. See also separate entry on **Hydraulic Controller.**

(*Impedance-Coupled Amplifier*)—An amplifier similar in form to a resistance-coupled amplifier in which the output voltage drop is developed across a choke or impedance coil. The choke or impedance coil can be designed to have a suitably high impedance at the frequency of the signal being applied and still have a resistance low enough so that only a relatively small voltage drop appears across it at dc. This form of coupling is usually used in amplifier design to amplify audio-frequency signals.

(*Intermediate-Frequency Amplifier*)—The amplifier used in superheterodyne receivers which amplifies the sum or difference frequency produced in the mixer or first detector by the heterodyning of the signal and oscillator frequencies.

(*Isolated Amplifier*)—See prior item in this list—Floating Amplifier.

(*Klystron Amplifier*)—A klystron tube may be used as an amplifier as well as oscillator in uhf applications. A cascade-amplifier klystron contains three resonant cavities for increased power amplification and output. The third resonator lies between the input and output resonators and has no external connection. It is excited by the bunched beam that emerges from the input-resonator gap, and it produces further bunching of the beam. See also **Microwave Tubes.**

(*Light Amplifier*)—The term laser stands for light amplification by stimulated emission of radiation. A laser is an active electron device that converts input power into a very narrow, intense beam of coherent light. See also separate entry on **Laser.**

(*Linear Amplifier*)—A pulse amplifier in which the output pulse height is proportional to an input pulse height for a given pulse shape up to a point at which the amplifier overloads.

(*Linear Power Amplifier*)—A power amplifier in which the signal output voltage is directly proportional to the signal input voltage.

(*Logarithmic Amplifier*)—An amplifier whose output signal is a logarithmic function of the input signal.

(*Magnetic Amplifier*)—A device that uses saturable reactors either alone or in combination with other circuit elements to achieve amplification. A simple form of magnetic amplifier is shown in Fig. 7. An alternating voltage source is connected in series with a load resistor R_L and two coils (called gate windings) having the same number of turns (N_g). The gate windings are wound on separate cores. A third winding (N_c turns), called a control winding, is wound around the two cores containing the gate windings. The magnetic amplifier functions as an amplifier in that a signal applied to the control winding controls the flow of current in R_L produced by the source of alternating voltage. The control is effected through the action of the gate windings which act as switches which are closed for a fraction of the alternating current cycle which depends on the amplitude of the signal applied to the control winding.

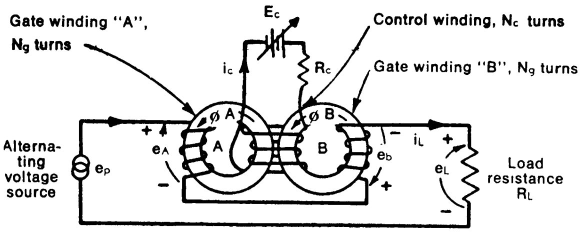


Fig. 7. Elementary magnetic amplifier.

The gate windings act as switches by virtue of the fact that their impedance is very high at one part of the cycle and very low at another part. The behavior of the impedance results from the characteristics of the magnetic material used for the cores on which the gate windings are located. The relation between the magnetomotive force produced by a coil wound on the core and the resulting magnetic flux in the core may be approximated by the graph shown in Fig. 8. Between $-M$ and $+M$ the flux curve has a very steep slope and outside of this range the graph is essentially a horizontal line. In the latter range the magnetic material is said to have reached saturation with a magnetic flux ϕ_s. The potential difference across an inductor is directly proportional to the rate of change of the magnetic flux intercepted by the coil. Conversely, the flux intercepted may be expressed as the time integral of the potential difference. If a sinusoidal voltage is impressed across the terminals of an inductor, the value of flux that results at all instants of time may be determined from the time integral of the sinusoidal voltage.

By knowing the form of the flux-magnetomotive force (ampere-turns) relation (Fig. 8, for example) and the number of turns in the coil, the current that flows through the coil to provide the necessary ampere turns can be determined. Examination of Fig. 8 shows that, for applied voltages such that the flux produced is less than ϕ_s, a comparatively small magnetomotive force is required and hence a relatively small current will result in the coil. The inductor has a very high impedance under this condition. If the applied voltage results in a flux greater than ϕ_s, a large value of magnetomotive force, and thus a large value of current, will be required to satisfy the flux condition. The impedance of the coil for voltage amplitudes above a certain amount will thus be very low. These notions can also be expressed in terms of the equivalent inductance of the coil which is proportional to the slope of the flux-magnetomotive force curve.

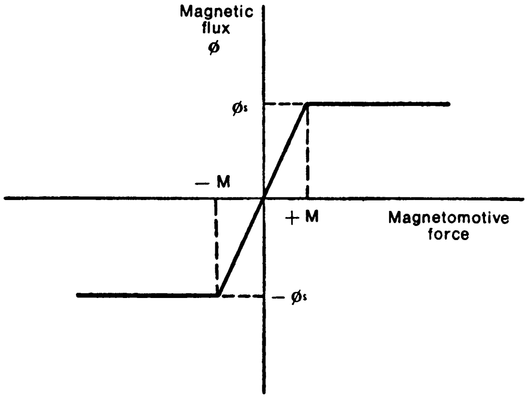


Fig. 8. Approximate relation between flux and magnetomotive force.

From Fig. 8 it is seen that, for applied voltages resulting in operation between $-M$ and $+M$, the inductance will be very large, whereas outside of this region the inductance will be essentially zero. If a sinusoidal voltage is applied to a coil wound upon a core having the properties shown in Fig. 8, then essentially no current will flow in the coil until the amplitude of the sine wave reaches a critical value (the value corresponding to a flux ϕ_s), but for the time that the amplitude exceeds the critical value, the voltage drop across the coil will be close to zero, and the current that results will be determined principally by the amplitude of the applied voltage and the impedance of any other elements connected in series with the coil.

The ratio of power gain to time constant can be improved if positive feedback is applied to the amplifier of Fig. 7. The circuit with the feedback connection, as well as a bridge rectifier to provide a unidirectional load current, is shown in Fig. 9. For simplicity the control winding is shown as two separate windings, one on each core. In the circuit, the load current which, although it pulsates, always flows in one direction, is forced to flow through two feedback windings which are arranged so

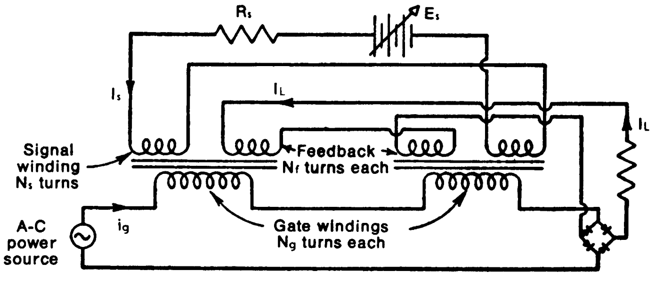


Fig. 9. Magnetic amplifier with external feedback.

that their magnetomotive forces reenforce those created by the control windings on both cores. The ampere turns contributed by control plus feedback windings are still approximately equal to those contributed by the gate windings, but because of the contribution of the feedback winding the control winding ampere turns for a given load current can be much less than without the feedback.

Two representative magnetic amplifier circuits are shown in Figs. 10 and 11.

(*Magnetron Amplifier*)—A traveling-wave magnetron used as an amplifier. The basic features of a typical structure are shown in Fig. 12. It resembles a section of a vane-type, cavity magnetron of infinite radius, excited at one end, and coupled to a load at the opposite end. A beam of electrons is projected through the space between the plane electrode and the cavity structure, which is maintained at a positive potential relative to the plane electrode. A magnetic field, normal to the plane of the paper, is adjusted so that the electrons do not strike the anode in the absence of alternating field. If the velocity of the electron beam is equal to, or nearly equal to, the phase velocity with which electromagnetic waves move down the loaded waveguide formed by the cathode and anode, energy is transferred to the electromagnetic wave from the source of direct voltage. The electromagnetic wave that moves from the input resonator to the output resonator, therefore, increases in amplitude, and the power output exceeds the power input.

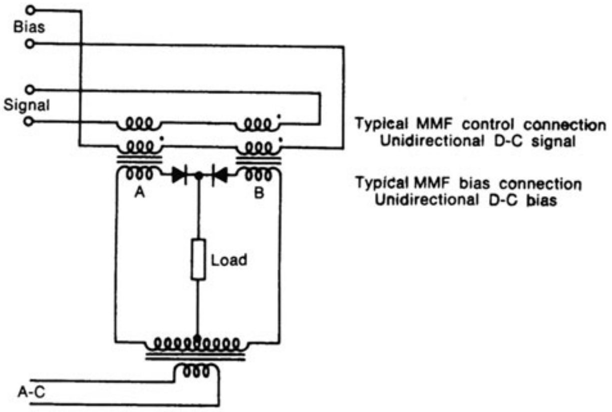


Fig. 10. Self-saturating magnetic amplifier circuit.

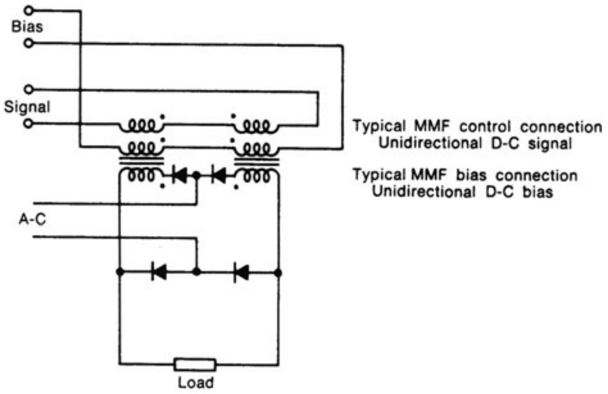


Fig. 11. Magnetic amplifier doubler circuit.

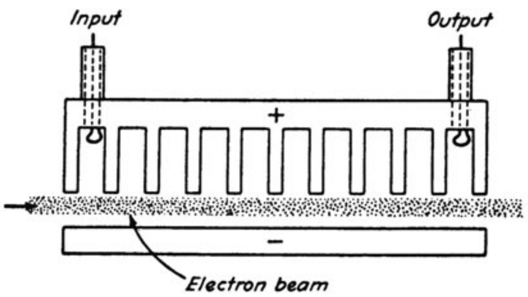


Fig. 12. Magnetron amplifier structure. Magnetic field is perpendicular to plane of paper.

(*Microwave Amplifier*)—The term maser stands for microwave amplification by stimulated emission of radiation. The amplification in a microwave amplifier is achieved by raising atoms or molecules of a paramagnetic material to an unstable high energy level. See also separate entry on **Maser.**

(*Modulated Amplifier*)—An amplifier stage in a transmitter in which the modulating signal is introduced and modulates the carrier.

(*Monitoring Amplifier*)—In broadcasting and recording, an amplifier with high input impedance and medium power output which is bridged across the program circuit. The available power output is used to operate a loudspeaker and/or headsets for the benefit of control personnel.

(*Nonlinear Amplifier*)—An amplifier in which the output is not related to the input by a simple constant. One form is the volume-limiting amplifier, where the average gain is changed in such a manner that steady-state waveforms are accurately reproduced; another form, frequently called a clipping or over-driven amplifier, has an output which is a greatly distorted version of the input.

(*Operational Amplifier*)—Used to perform analog-computer functions, an operational amplifier is an amplifier with high dc stability and high immunity to oscillation, usually achieved by using a large amount of negative feedback. The operational-amplifier integrator shown in Fig. 13 is used in the conversion of voltage to a frequency. This is an amplifier with a capacitor connected between output and input and is used in digital voltmeters. As shown, there is a resistor in front of the input to the integrator. The amplifier is assumed to have a high gain. As the result of this configuration, the application of a step input results in a linear ramp output in which the slope of the ramp is proportional to the input-voltage step. The basic circuit of another integrating device making use of an operational amplifier is shown in Fig. 14. The voltage appearing at the output of the amplifier e_0 is proportional to the integral of the input voltage e_i, or $e_0 = 1/RC \int_{t_0}^{t_1} e_i \, dt$.

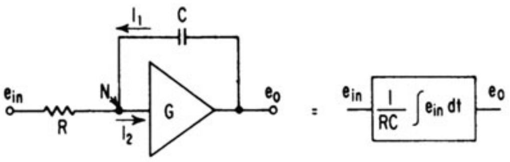


Fig. 13. Operational-amplifier integrator.

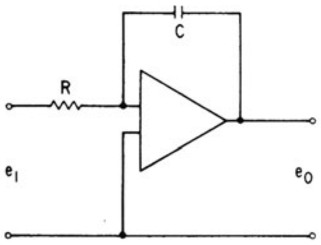


Fig. 14. Operational amplifier with capacitance feedback

(*Paramagnetic Amplifier*)—An amplifier which increases the power level of a signal by means of the variations of an energy-storage parameter.

The classic example is the L-C resonant circuit in which the spacing of the capacitor plates is varied cyclically by a "pump" at a frequency ω_p, twice the signal frequency ω_s, which is identical to the resonant frequency of the circuit. Initial charge on the capacitor will result in a sinusoidal variation of capacitance charge as a function of time. If the capacitance is always decreased by the pump at the plus or minus peak of charge during which time charge is relatively constant, an increase in capacitor voltage must result, since $C = QV$. This increase in capacitor voltage represents an increase in the power level of the original signal, the additional energy representing the work done by the pump in increasing the separation of the capacitor plates against the forces of electrostatic attraction.

Maximum energy will be transferred from the pump frequency to the signal frequency if the pump frequency is twice the signal frequency and if the relative phase of the two is adjusted as described above. In a practical case, the separation of the two frequencies cannot be conveniently maintained constant. Therefore, the signal and pump will no longer interact favorably all the time, but will drift periodically into and out of the optimum condition. This produces a modulation of the signal frequency, resulting in frequency components in the output, among which are ω_s, ω_p, and ($\omega_p - \omega_s$) the latter term being called the *image* or *idler* signal.

The mode of operation described above is called a *negative resistance amplifier*, since it acts to neutralize or overcome the positive resistance of the resonant circuit. Stability problems arise, since the output and input terminals are inherently the same. The closeness of the image frequency to the signal frequency makes the separation of the two difficult, one procedure being the utilization of a circulator. The pump signal may be filtered from the output by a conventional band-elimination filter.

A slightly different mode of operation, called variously the *up-conversion amplifier* or *amplifying-up converter*, differs from the negative resistance type of amplifier in these respects: (1) Since the gain can be shown to be proportional to $\omega_s + \omega_p/\omega_s$, the pump frequency is made as high as is practicable. (2) An additional frequency of interest is ($\omega_s + \omega_p$). The signal frequency is inevitably shifted in the amplification process. (3) The up-converter is a two-part device with unconditional stability, without the requirement of a circulator.

A reverse-biased semiconductor diode provides an excellent nonlinear reactive element for use in a parametric amplifier inasmuch as it presents a very high shunt resistance and a capacitance which varies approximately inversely with the square root of the applied voltage. Thus, it is common practice to employ these elements as the capacitor C in Fig. 15. Parametric amplification can also be achieved by the utilization of the nonlinear characteristics of ferromagnetic and ferroelectric materials.

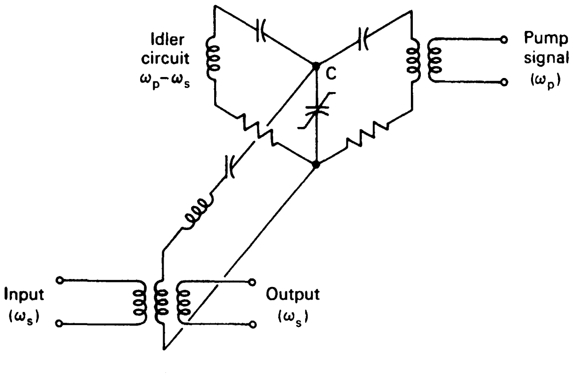


Fig. 15. Parametric amplifier.

(*Paraphase Amplifier*)—A phase-inverter amplifier used to convert a single-ended signal to a push-pull signal.

(*Pentriode Amplifier*)—A video amplifier containing a pentode which by virtue of suitable bypass and coupling devices, is made to operate as a triode (screen effectively connected to the plate) over a portion of the frequency range, and as a pentode (screen-grounded) over another part of the frequency range.

(*Pneumatic Amplifier*)—A device that increases signal strength pneumatically—as say a 3–15 psi (155–776 torr) instrument signal to several hundred psi to operate a large piece of process equipment, such as a slide valve, damper, furnace door, etc. Usually, such terms as booster, positioner, and relay, are used instead of amplifier.

(*Power Amplifier*)—An amplifier adjusted to operate with the objective of developing power in the load impedance in contrast with emphasis on production of output voltage or current. In many applications it is necessary to get power from an amplifier unit, so at least the final stage is usually adjusted to give power rather than voltage output. In radio-transmitting circuits or large power audio-amplifiers several stages may be power amplifiers. For this type of service the circuit is adjusted to give as large a current as possible through the load, which has much lower resistance than in voltage amplifiers. The transistors used are somewhat larger as a rule than those used for voltage and current amplification and are specially designed for large current outputs.

(*Preamplifier*)—This is a voltage amplifier, which receives the signal from a microphone, pick-up, television camera tube, or other device supplying a low signal level and amplifies it so it can supply the input for additional amplifier circuits. Thus a preamplifier is commonly used in a radio or television studio to amplify the audio or video signal before feeding it into a mixer, line to the transmitter, or other amplifying equipment at the studio.

(*Program Amplifier*)—The amplifier following the master volume control in an audio console. Its gain brings the signal to a level suitable for transmission.

(*Push-Pull Amplifier*)—An amplifier in which the input signals applied to the transistors are opposite in phase and the output signals are combined to obtain twice the output of a single stage. Push-pull amplifiers have lower harmonic distortion than amplifiers with one transistor and with transformer-coupling they have the added advantage that the collector currents for the two devices act to magnetize the transformer core in opposite directions, tending to avoid saturation of the core. If the devices are perfectly balanced, no signal voltage exists across the bias resistor and a bypass capacitor normally used can be omitted.

The property of complementary symmetry makes it possible to dispense with a phase inverter in some push-pull transistor amplifiers. Figure 16 shows a push-pull amplifier formed from two common-collector connected transistors. Since the circuit has a low output impedance, it may be used for direct coupling to a loudspeaker without a transformer. This arrangement employs one *n-p-n* and one *p-n-p* transistor. Because of the complementary symmetry property, a positive signal current injected into the base of the *n-p-n* unit will cause its collector current to increase, whereas the same positive injected signal will cause the collector current of the *p-n-p* unit to decrease. As a result, the two base connections may be supplied from one signal source. In the figure the resistors connecting collectors and bases provide a small forward bias on the transistors. The resistance R is sufficiently low so that both transistors have essentially equal input currents.

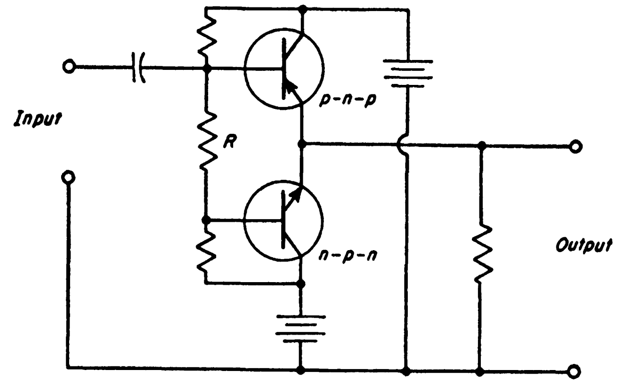


Fig. 16. Complementary symmetry push-pull amplifier.

(*Reflex Circuit Amplifier*)—A circuit in which a transistor simultaneously amplifies signals in two widely-separate frequency bands (i.e., the intermediate frequency signal of a superheterodyne and the audio-frequency output of the detector). Used very rarely because the savings of material and space are not justified by the additional complexity of operation.

(*Sample-and-Hold Amplifier*)—Also known as a track-and-hold amplifier, this device has an output that is proportional to the input until a "hold" signal is received. Upon receipt of that signal, the amplifier output is maintained essentially constant even though there may be changes in the input signal. See also separate entry on **Sample-and-Hold Amplifier.**

(*Servoamplifier*)—An amplifier used in a control or servosystem.

(*Single-Ended Amplifier*)—An amplifier in which each stage normally employs only one transistor, or if more than one such device is used, in which they are connected in parallel so that operation is asymmetric with respect to ground. See also separate entry on **Single-Ended Amplifier.**

(*Single-Ended Push-Pull Amplifier*)—A form of amplifier in which a pair of output terminals may have an instantaneous voltage of either polarity as may be dictated by the phase of the input signal.

(*Stagger-Tuned Amplifier*)—An amplifier incorporating staggered tuning to provide a desired bandwidth characteristic.

(*Step-Down Amplifier*)—A vacuum-tube type amplifier used to measure very high potentials which are impressed between anode and cathode, the anode being negative. The corresponding grid current is measured with the grid positive. This is sometimes called an "inverted voltmeter."

(*Transformer-Coupled Amplifier*)—An amplifier in which the interstage coupling from the output of one amplifying stage to the input of the next or to the load impedance is made using a transformer. The coupling method is used in audio-frequency amplifiers. The method is particularly advantageous in the opportunity offered for impedance matching when the load is a speaker or a low impedance transmission line. In Fig. 17 is shown the audio output stage of a transistor radio receiver. Changes in the base current resulting from the input signal cause corresponding changes in collector current which flows through the primary winding of the transformer. There results an alternating voltage across the secondary which has the same form as the input signal to the base and is capable of providing the necessary power output from the speaker. In this instance, the transformer makes possible matching the relatively low impedance of the speaker to the much higher output impedance of the transistor.

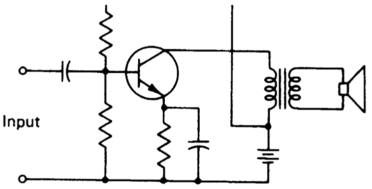


Fig. 17. Transistor audio amplifier.

(*Traveling-Wave Amplifier*)—An amplifier that uses one or more traveling-wave tubes to provide amplification of signals at frequencies of the order of thousands of megacycles.

(*Tuned Amplifier*)—An amplifier in which the load impedance consists of, generally, a parallel inductance-capacitance network, or two or more of these having electromagnetic coupling. The fact that the impedance of this network varies with frequency causes the gain of the amplifier to vary as a function of frequency in a somewhat similar manner. A tuned amplifier, employing a transistor and used for the amplification of intermediate frequency signals, is shown in Fig. 18. A tuned primary circuit is used and the transistor is connected across only a portion of the inductor. The arrangement offers advantages from the standpoint of impedance matching which permit higher gains to be achieved than would be possible otherwise in view of typical transistor characteristics.

(*Video-Frequency Amplifier*)—A device capable of amplifying such signals of wide bandwidth as are used in television and radar.

(*Voltage Amplifier*)—An amplifier designed for the primary purpose of producing an increase in signal voltage with little or no attention to the available output power of the stage.

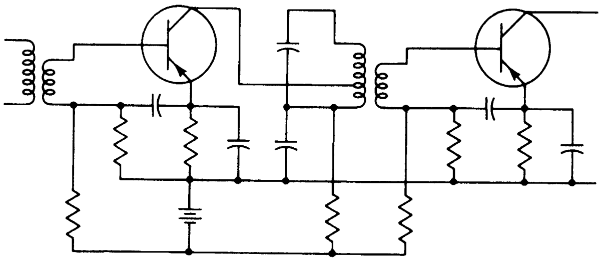


Fig. 18. Transistor intermediate-frequency amplifier.

(*Volume-Limiting Amplifier*)—An amplifier containing an automatic device which functions when the input volume exceeds a predetermined level, and so reduces the gain that the output volume is thereafter maintained substantially constant, notwithstanding further increase in the input volume. The normal gain of the amplifier is restored when the input volume returns below the predetermined limiting level.

(*Wide-Band Amplifier*)—An amplifier having uniform response over many decades of frequency. An example is the video amplifier.

Additional Reading

Belove, C.; Ed.: "Handbook of Modern Electronics and Electrical Engineering," John Wiley & Sons, New York, 1986.

Fink, D. G., and D. Christiansen, Eds.: "Electronics Engineers' Handbook," McGraw-Hill, New York, 1982.

Freeman, R. L.: "Reference Manual for Telecommunications Engineering," John Wiley & Sons, New York, 1985.

Grob, B.: "Basic Electronics," 5th Ed., McGraw-Hill, New York, 1984.

Kaufman, M., and A. H. Seldman: "Handbook for Electronics Engineering Technicians," 2nd Ed., McGraw-Hill, New York, 1986.

Pasahow, E.: "Electronics Ready Reference Manual," McGraw-Hill, New York, 1986.

Stout, D. F., and M. Kaufman: "Handbook of Operational Amplifier Circuit Design," McGraw-Hill, New York, 1976.

AMPLITUDE COMPARISON. The process of indicating the time at which two waveforms reach the same amplitude. It may also be considered to be the method of determining the abscissa of a waveform, given its ordinate.

AMPLITUDE DISCRIMINATOR. A circuit which performs an amplitude comparison. In addition, the sense and magnitude of the inequality of the amplitudes may be obtained.

AMPLITUDE DISTORTION. A type of distortion that occurs in an amplifier or other device when the amplitude of the output is not exactly a linear function of the input amplitude.

AMPLITUDE MODULATION. See **Modulation.**

AMPLITUDE SEPARATION. The process of separating all values of a wave greater or less than a given amplitude, or those lying between two amplitudes.

AMPOULE. Sometimes spelled ampule, a small sealed glass container for drugs that are to be given by injection. As they are completely sealed, the contents are kept in their original sterile condition.

AMPULLA. Any flask-like dilatation, such as the small saccular outgrowth of the water vascular system of starfishes, at the inner end of the tube foot, or the dilated portion of the semicircular canals of the vertebrate ear.

AMYGDALOID. A vesicular rock, commonly a lava, whose cavities have become filled with a secondary deposit of mineral material such as quartz, calcite, and zeolites. The term is derived from the Greek word meaning almond in reference to the frequent almondlike appearance of the filled vesicles which are called amygdales or amygdules.

ANABOLISM. See **Basal Metabolism.**

ANACLINAL. Used in structural geology to define a direction opposite to the dip of the strata or formations.

ANAEROBE. An organism that can grow in the absence of free oxygen and referred to as an anaerobic organism. Subdivided into *facultative anaerobes*, which can grow and utilize oxygen when it is present; and *obligatory* anaerobes, which cannot tolerate even a trace of oxygen in their surroundings. Yeast is an example of a facultative anaerobe. Yeast can grow and utilize oxygen in its metabolism, in which case it utilizes all the energy in a carbohydrate and yields water and carbon dioxide. In the absence of oxygen, the yeast cells turn to fermentation and anaerobic metabolism wherein the carbohydrate is converted into alcohol and carbon dioxide. The bacterium which produces botulism in "preserved" foods is an obligatory anaerobe (*Clostridium botulinum*).

In addition to botulism, other species of Clostridium are the etiological agents of tetanus, gas gangrene, and a variety of animal diseases including struck, lamb dysentery, pulpy kidney, and enterotoxemia. The pathogenicity of the clostridia depends largely on the production of potent toxins. Generally, these are heat labile exotoxins and often several pharmacologically and immunologically different toxins may be produced by the same species. In recent years, certain of the toxins have been purified to the point of crystalline purity and revealed to be high-molecular-weight proteins. The availability of the pure toxins has permitted more precise studies both of the chemical nature of the toxin and also of their properties as enzymes. From such toxins, it is possible to produce efficient toxoids for use in prophylactic immunization measures for tetanus and botulism. Some clostridia, in contrast, are beneficial. In addition to the nitrogen-fixing ability of certain species, *C. sporogenes* is active in the decomposition of proteins and contributes to biodegradability of substances, particularly of plant remains. *C. acetobutyleium* is active in production of acetone and butyl alcohol by fermentation.

Glucose can be broken down through glycolysis into pyruvic acid, but in the absence of oxygen as a final acceptor for the hydrogen, the pyruvic acid cannot enter the tricarboxylic acid cycle for further breakdown. Instead the pyruvic acid itself serves as an acceptor for the hydrogen split off in glycolysis. Hence, much less energy is obtained from food in anaerobic metabolism. Lactic acid, alcohol, and some extremely poisonous substances are products of anaerobic metabolism.

In addition to the one-celled anaerobes, certain parasitic worms, such as *Ascaris*, are thought to use anaerobic metabolism, to a certain extent at least. Living in the intestine of higher animals these worms have little access to free oxygen.

The muscles of higher animals are also known to use anaerobic metabolism when the demands on the muscles for energy are greater than can be supplied through the available oxygen. Lactic acid is generated.

The diverse mechanisms that permit animal life in the absence of oxygen are described in considerable detail by P. W. Hochachka ("Living without Oxygen," Harvard Univ. Press, Cambridge, Massachusetts, 1980).

ANALCIME. A common zeolite mineral, $NaAlSi_2O_6 \cdot H_2O$, a hydrous soda-aluminum silicate. It crystallizes in the isometric system, hardness, 5–5.5; specific gravity, 2.2; vitreous luster; colorless to white; but may be grayish, greenish, yellowish, or reddish. Its trapezohedral crystal resembles garnet but is softer; it is distinguished from leucite only by chemical tests.

There are many excellent European localities. Magnificent crystals occur at Mt. St. Hilaire, Quebec, Canada; in the United States at Bergen Hill and West Paterson, New Jersey, Keweenaw County, Michigan, and Jefferson County, Colorado. Nova Scotia furnishes beautiful specimens.

Analcime is a relatively common mineral and occurs with other zeolites in cavities and fissures in basic igneous rocks, occasionally in granites or gneisses. It seems to occur as a replacement and perhaps in some cases as a primary mineral crystallizing from a magma rich in soda and water vapor under pressure. The name analcime is derived from the Greek word meaning weak, in reference to the weak electric charge developed when heated or subjected to friction.

ANAL FEELERS. Posterior sensory appendages such as the anal cirri of annelid worms and the cerri of insects.

ANALGESICS. Drugs which diminish sensitivity to pain without impairing consciousness. These drugs act on the central nervous system and include opiates, coal tar analgesics, aminopyrine, salicylates, and phenylbutazone. Some of the analgesics also act in other ways pharmacologically. In the case of phenylbutazone, the excretion of uric acid is promoted. Salicylates reduce fever. Of the various alkaloids, only morphine and codeine are analgesics. Although still a valuable drug for some situations, particularly for relieving intense pain, morphine has several objectionable characteristics, including its toxicity and addicting nature. Codeine is a much less powerful drug and with much less objectionable aftereffects. See also **Alkaloids;** and **Morphine.**

Colchicine, an alkaloid, is used for the abatement of swelling and pain in acute attacks of gout. Normally, colchicine is not considered an analgesic, but in this circumstance, it does play this role as well as serving as an antipyretic and antiphlogistic (counteraction of fever and inflammation).

For the relief of ordinary aches and pains, the coal-tar analgesics are commonly used. These include acetanilid, acetophenetidin (phenacetin), and N-acetyl-*p*-aminophenol. They frequently are mixed with caffeine and aspirin, or a barbiturate. Acetanilid is somewhat more toxic than acetophenetidin and thus is used less frequently. A side effect of N-acetyl-*p*-aminophenol may be minor gastrointestinal distress.

Aminopyrine, an analgesic and antipyretic, is used less frequently because it can cause agranulocytosis. See also **Agranulocytosis.**

Of the salicylate drugs (derivates of salicylic acid), sodium salicylate and acetylsalicylic acid (aspirin) are the most widely used. The latter is poorly soluble in water and hydrolyzes into salicylic and acetic acids. Aspirin is combined with such compounds as phenacetin (acetophenetidin) and caffeine in a number of proprietary preparations. Aspirin is the most widely used medicine in the United States. Generally, aspirin is much more effective for pain originating in joints and muscles than in the internal organs. Aspirin also performs well in the treatment of acute rheumatic fever and is a uricosuric drug (stimulates excretion of uric acid). The latter causes symptoms of gout. Large quantities of aspirin also are consumed by arthritics. Aspirin is not totally harmless even though used widely. It can cause allergic reactions among some individuals. See also **Acetylsalicylic Acid.**

ANALOG COMPUTER. A computer that solves problems by physical analogy. The computer translates temperature, flow, speed, altitude, voltage, and other physical variables into related electrical quantities and uses electrical equivalent circuits as an analog for the physical phenomenon being investigated. On an analog computer, for example, physical characteristics, such as weight or temperature are represented by voltage. Voltage is the electrical analog of the variable that is being analyzed. The variable itself can be electrical as well as hydraulic or pneumatic. Scale factors relate voltages in the computer to the variable in the problem being solved.

Components of the computer may be designed to operate within any fixed output voltage range. However, only two ranges are in wide use: ±100 volts and ±10 volts. All computer variables are scaled to lie within one of these ranges. For example, a temperature which varies from 0 to 1000°C is represented on a 100-volt computer by a voltage varying from O to 100. The scale factor would be 1/10 volt per °C.

Although analog computers still find application for certain needs, the practice for several years has distinctly reflected a preference for digital computers. For example, the slide rule (an analog computer) was for many decades the "right hand" of engineers and scientists. Although still used on occasion, the slide rule essentially has been displaced by desk calculators, pocket calculators, and personal computers.

ANALOG INPUT. This term is used to describe an assembly of equipment (subsystem) for performing the selection of the analog signal to be sampled, modification or conditioning of the analog signal, analog-to-digital conversion to provide a digital representation, and the necessary digital controls for subsystem control and communication with a digital computer. Generally included in this assembly are signal conditioners, such as attenuators and filters, analog signal multiplexers, amplifiers, analog-to-digital converters, and the control logic. Check alphabetical index.

For data-acquisition or process-control computers, there are two fundamental analog-input subsystem types. However, within each of these basic configurations, there are numerous variations possible as dictated by the use of different kinds of multiplexers, amplifiers, and analog-to-digital converters. Cost and performance are the usual predominating factors in selection. The two major subsystem types are: (1) high-level, and (2) low-level systems.

High-Level Analog-Input Subsystems: High-level signals typically are in the range of 0 to ± 10 volts. These are connected to the subsystem input terminals of a configuration of the type shown in Fig. 1. Including filtering, limiting to protect the analog-input subsystem from overvoltage, and attenuation, the signal conditioning of signals is performed on a *per-channel* basis. The multiplexer selects the signal to be converted and is under the control of the subsystem control logic. Isolation between the analog-to-digital converter and the input-signal source is provided by a buffer amplifier. A sample-and-hold amplifier may be used to reduce aperture-time errors. The buffer amplifier sometimes is included as an integral part of the analog-to-digital converter. The timing and sequencing of the subsystem operation is transmitted by the control logic, which also transmits the digital value to the digital computer.

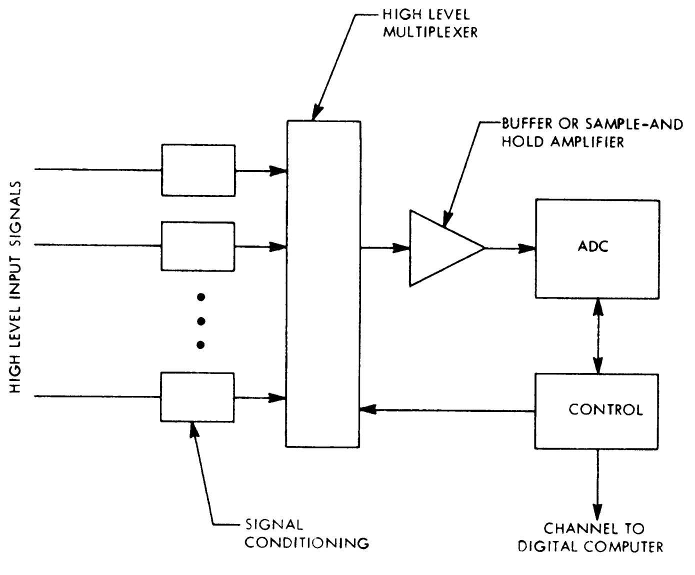


Fig. 1. Schematic diagram of high-level analog-input subsystem.

Usually configurations of these subsystems are single-ended, although some differential systems are obtainable. The latter are more costly, but offer easier installation because noise and ground loops are less serious. Reduction of the latter problems is important particularly in the case of high-speed, high-resolution applications and where input signals must come from signal sources a considerable distance from the subsystem.

Solid-state switches, capable of sampling signals at very high rates, normally are used in the analog multiplexer. Both bipolar transistors and field-effect transistors are used. Bipolar transistor switches restrict levels to a high level because of voltage-offset errors. The most common analog-to-digital converter used is of the successive-approximation type, since its speed matches the solid-state multiplexer characteristics. For lower-speed applications, ramp and integrating-ramp analog-to-digital converters can be used.

Timing and channel communications functions are provided by the control logic. The digital computer normally initiates a control word, specifying one or more input addresses, to commence operation of the subsystem. The control unit selects the appropriate multiplexer switches. After a predetermined delay time, the control unit provides a "start convert" signal to the analog-to-digital converter. The latter operates asynchronously with respect to the computer and, at the end of the conversion, signals the control logic. Prior to transmitting the result back to the digital computer, the control unit may reformat the output word from the analog-to-digital converter, that is, adding parity bits or converting from parallel to a serial form. It may also collect some or all of the results in a buffer storage where they are available to the computer on demand.

The characteristics of the sampling subsystem are determined by the type of multiplexer, amplifier, and analog-to-digital converter used. Under controlled operating conditions, a total measurement error of less than 0.05% of full scale is desirable.

Subsystems of the type just described have been used in numerous high-speed data-acquisition applications, including hybrid computation and rocket test stand monitoring, as well as for general research. These uses typically have wide channel-bandwidth requirements, high-level signals, and/or a relatively small number of signals. Because the configuration is not adapted to the handling of low-level signals, it is not used widely for process control. For applications involving only a few data points, a high-speed low-level system can be achieved by adding a high-gain amplifier for each input signal.

Low-Level Analog-Input Subsystems: Systems of this type are similar to the high-level system just described with the exception of the use of a time-shared low-level amplifier. See Fig. 2. In operation, the input signals are conditioned and multiplexed by a low-level multiplexer. Before conversion into a digital representation by the analog-to-digital converter, the multiplexer output is amplified by a time-shared amplifier. Measurement accuracy for the total system for a 50-millivolt signal range typically is 0.1%, with a resolution of 10 to 14 bits.

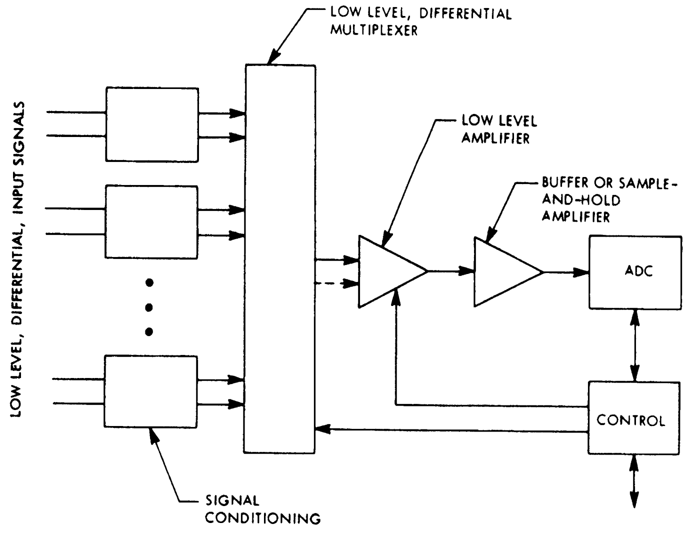


Fig. 2. Schematic diagram of low-level analog-input subsystem.

The low-level signals generally are less than 100 millivolts and often as low as 10 mV full scale. Thus, this type of subsystem usually provides a differential-signal input. See also **Differential-Mode Voltage.** The multiplexer and signal-conditioning circuits are differential. The amplifier may or may not be differential, depending on the type of multiplexer used. A single-ended amplifier may be used where the multiplexer provides common-mode isolation, such as a transformer-coupled multiplexer. The output of the amplifier and the analog-to-digital converter input are single-ended in most cases.

Differential-multiplexer configurations for low-level uses most commonly are the double-pole single-throw differential, flying-capacitor, and transformer-coupled designs. Switching devices normally are dry-reed relays, field-effect transistors, or mercury-wetted contact relays. Field-effect transistor switches can be applied for very high speeds. Electromechanical devices are limited to about 300 samples/second or less. In any case, however, subsystem speed normally is limited to less than 10,000 samples/second because of amplifier performance. With the proper selection of multiplexer configuration and the capabilities of the amplifier, common-mode voltages up to several hundred volts will not affect the acceptable performance of the system. However, where solid-state switches are used in multiplexers, the tolerance is usually limited to something less than 30 volts of common-mode voltage.

In the design of Fig. 2, a differential or single-ended amplifier may be used, depending on the multiplexer characteristics. In some cases, the amplifier will have a fixed gain. In other instances, the control logic may select the gain, either automatically or under control of the computer program. Where a differential amplifier is used, the common-mode tolerance of the amplifier is critical to determining the common-mode rejection of the subsystem. Low-level applications also require that amplifier noise be kept to a minimum.

The design configuration just described has been used in process control and large data-acquisition systems. Such uses typically have low channel-bandwidth requirements and a large number of low-level signals. The time-shared amplifier provides the system with a cost advantage as compared with an amplifier-per-channel type system. The low band-width application characteristic makes it possible to use the limited system-sampling speed in numerous applications.

Thomas J. Harrison, International Business Machines
Corporation, Boca Raton, Florida.

ANALOG MULTIPLEXER. An array of analog switches used for the selection of one of several analog signals for transmission to subsequent devices in a data acquisition subsystem. Multiplexers most typically are used for signal selection prior to amplification and conversion to digital form in time-shared subsystems. An analog multiplexer used in the analog-input subsystem of a data-acquisition system is shown schematically in Fig. 1. Multiplexers are not confined to data-acquisition systems, but are used in audio-switching networks in telephony. The particular hardware configuration varies considerably with the nature of the signals and their manipulation. See Fig. 2.

The control unit supervises switch selection. Where the multiplexer is differential, two or three switches are required for each input signal. Where the multiplexer is single-ended, only one switch per input is required, using a common ground to complete the circuit for all input signals. Although the control unit varies with particular applications, its primary function is to furnish the signals required to select or address the various switches in the multiplexer at the proper time. The logic in the control unit or a command from a computer will determine the particular address to be selected. In some cases, the switches may be selected sequentially under control of a ring counter circuit in the control unit. Or, the selection of each address may require a data word from the computer or microprogrammed control unit.

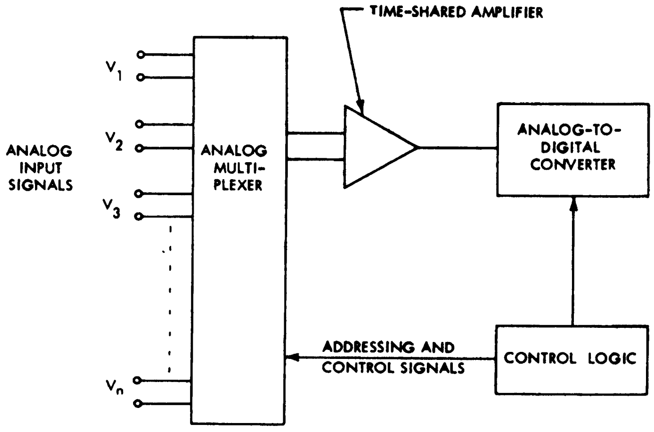


Fig. 1. Location of analog multiplexer in analog input system.

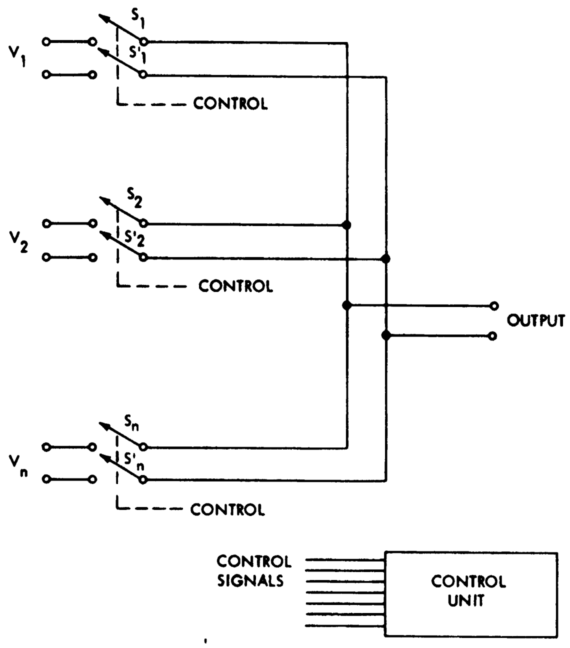


Fig. 2. Generalized configuration of analog multiplexer.

The control unit also furnishes timing for the multiplexer and associated equipment. Usually, the analog-to-digital (A/D) converter cannot commence a conversion instantly after selection of the multiplexer switch. A time delay is required to permit actuation time of the switch, for settling of the amplifier, and for the dissipation of other transients. Counters or delay circuits in the control unit usually furnish the required timing signals.

Several types of solid-state or electromechanical analog switches may be used in the multiplexer. Types of electromechanical switches used are mercury-wetted contact relays, dry-reed relays, and crossbar switch assemblies. Of the solid-state switches, most often field-effect transistors are used. For special applications, diodes, bipolar transistors, and silicon-controlled rectifiers may be used.

Among the major performance characteristics of an analog multiplexer are accuracy, sampling speed, noise. Common-mode rejection ratio also is important in the instance of a differential multiplexer. The features of the multiplexer switching device mainly determine sampling speed. The electromechanical devices, such as mercury-wetted contact relays and dry-reed relays, usually are limited to about 250 samples/second. Bipolar and field-effect transistors can provide very high sampling speeds. Although a switch may close rapidly, a time interval must be provided to permit transients to dissipate prior to conversion of the multiplexer output signal to digital form. In the case of the mercury-wetted relay, the actual switch-closing time is in the range of 1 millisecond, but there may be a noise transient which will persist at levels in excess of 10 microvolts for perhaps 5 milliseconds or even longer. Where a noise level of this magnitude cannot be tolerated, a delay is required between switch selection and conversion initiation. Noise considerations also effect solid-state switches, but noise duration times are generally much shorter.

Careful consideration must be given to possible errors that a multiplexer can contribute to the operation of a digital-data acquisition subsystem. Leakage currents associated with the "off" switches will, in the case of solid-state switches, flow through the source impedance of the channel being sampled by the "on" switch and also through the multiplexer load impedance. Offset in the voltage being sampled thus can be caused by these currents. Even though leakage current from a single switch may be small (considerably less than a microampere), an appreciable error can result from the cumulative effect of all the switches that are connected to the common multiplexer output bus.

Inasmuch as the field-effect transistor does not show an offset voltage in the "on" condition, it may be used for multiplexing low-level signals. It should be noted, however, that some field-effect transistors show a significant "on" resistance and, because of source loading effects, can introduce errors. Further, the "on" resistance increases the effect of leakage currents as previously described. This may increase the output-signal rise time.

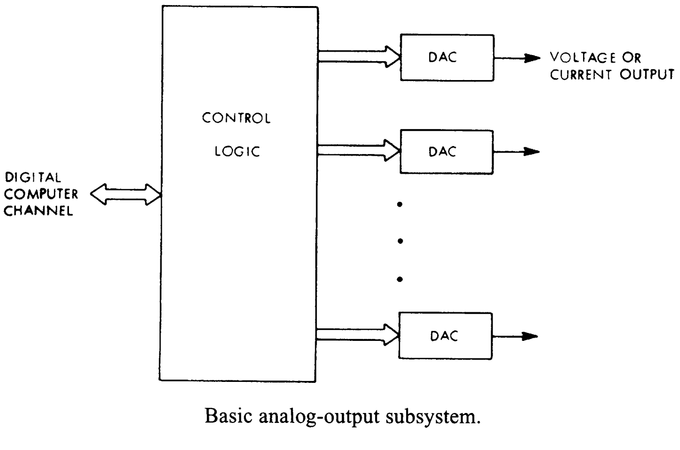
Where differential multiplexers are used, the common-mode rejection ratio is important. This factor is determined by leakage impedances between the switch-signal path and the drive circuit. Further, the maximum common-mode voltage which may be applied without causing damage to the multiplexer is determined by the breakdown voltage of the switch. Electromechanical devices have high drive-to-signal path and contact-breakdown voltages. Hence, these devices can be used for multiplexers that will withstand several hundreds of volts of common-mode voltage. Generally, solid-state multiplexers are limited to a common-mode voltage less than 20 to 30 volts. Improvement can be obtained through the use of isolation techniques, such as transformers.

The chances of multiplexer errors increase as the number of input channels in the multiplexer increases. Additional channels decrease the common-mode rejection ratio and also increase the offset resulting from leakage currents. *Submultiplexing* or *block switching* are techniques frequently used to minimize these conditions. Block switching entails the use of two levels of multiplexing. The first level is the same as that indicated in Fig. 2. This level furnishes selection of an input signal. The second level of multiplexing is provided at the output of the first-level multiplexer. As an example, a second-level switch may be provided for each group of 16 first-level multiplexing switches. Of course, in the selection of an input, both the first-level switches and their associated second-level switch must be actuated. The advantage of second-level multiplexing is the isolation of each block from the leakage currents and other possible disturbances which may arise from other inputs. In a system of 16 blocks of 16 channels each, for example, the leakage errors are determined mainly by the 15 "off" switches in the same block as the addressed point and the 15 "off" block or second-level switches. Hence, these errors equal those encountered in a 30-channel single-level multiplexer although a total of 256 input channels are serviced through the two-level multiplexer.

Thomas J. Harrison, International Business Machines Corporation, Boca Raton, Florida.

ANALOG OUTPUT. This term is used to describe an assembly of equipment (subsystem) and operations in a process control computer or data-acquisition system with the capability to provide a continuous voltage or current output which can be controlled by a digital computer. Closed-loop control systems, for example, which utilize a digital process control computer require current signals which may have ranges of 4 to 20, 1 to 5, or 10 to 50 milliamperes. These signals, in turn, are used for controlling process actuators, such as valve positioners. For the generation of visual displays, recorders and cathode-ray tubes require similar analog signals.

The digital-to-analog (D/A) converter is the principal component of the analog-output subsystem. Simply defined, a D/A converter is an electronically controlled attenuator network and a constant reference source. Digital-input signals received by the D/A converter activate analog switches which determines the attenuation factor of a passive network. Either electromechanical or solid-state switches may be used. The input to the attenuator network is a constant-voltage or -current source. The output of the network is proportional to the attenuator switch settings and, thus, to the digital-input signals.



Basic analog-output subsystem.

Connection of the D/A converter to the digital computer is through control logic. The latter provides addressing for each D/A converter and also the required control and timing signals. Shown in the accompanying diagram is an analog-output system of the kind used in process control computer or data acquisition systems. The digital computer, via digital-control words, specifies which D/A converter is to be adjusted. Another digital data word carries the desired output value. The address information is decoded in the control logic. The digital data also are routed to the input of the addressed D/A converter. This setting of the input switches of the D/A converter provides the desired analog-output value. The kind of subsystem shown is frequently used where the analog-output value is changed often, at a rate in excess of 1000 samples/second. This subsystem configuration can adjust many D/A converters at rates exceeding 100,000 samples/second. The resolution of the D/A converters used for applications of this type usually ranges from 8 to 13 bits. The total error of the output voltage, under controlled operating conditions, may be less than 0.01%.

Thomas J. Harrison, International Business Machines Corporation, Boca Raton, Florida.

ANALOG SWITCH. This term applies to a large family of switches that are designed for switching analog signals and normally implies high accuracy and high resolution. In data-acquisition and instrumentation systems, the most important switch characteristics are speed, voltage and current errors, "on" and "off" resistance, and noise.

Electromechanical analog switches, including mercury-wetted contact and dry-reed relays, usually have a lower "on" resistance and a higher "off" resistance than most solid-state switches. Also, because of excellent isolation between drive and signal circuits, electromechanical switches have no inherent voltage offset and leakage currents. Disadvantages include their slow performance as compared with solid-state switches and the production of noise by the switching action, a condition which tends to persist for an appreciable period.

Solid-state switches on the other hand are high-speed devices and with "on" and "off" resistances, compared with electromechanical switches, as described above. Where a *pn* junction is part of the signal path, as in a bipolar transistor, the switch will show an inherent voltage offset. The leakage currents between drive and signal paths also are larger in bipolar transistors as the result of less isolation between drive and signal. Field-effect transistors display leakages comparable with those of electromechanical switches. Noise results from switching action mainly because of coupling of the drive signal into the signal path by way of interelectrode capacitance. Even though the noise magnitude may be appreciable, decay to a negligible value occurs much more rapidly than in electromechanical switches.

Use of the major classes of analog switches is summarized as follows:

Relays and Other Electromechanical Devices—used for low-level signals (less than 1 V full-scale) and multiplexing speeds generally not exceeding 250 samples/second.

Field-Effect Transistors—used for low-level applications at higher speeds. Limited in common-mode-voltage handling capability as result of lower breakdown voltages.

Bipolar Transistors—used for high-level signals, but with appropriate circuitry for compensation of offset voltages; also may be used for low-level signals. They have been used less frequently in recent years due to improvements in field-effect transistor technology.

Thomas J. Harrison, International Business Machines Corporation, Boca Raton, Florida.

ANALOG-TO-DIGITAL CONVERTER. Abbreviated A/D converter or ADC. A device that provides a digital representation of an analog quantity. Examples of the latter include voltage, current, or a position. There are two principal types of A/D converters used in data-acquisition systems: (1) electromechanical, and (2) electronic converters. The electromechanical types sometimes are referred to as shaft- or position-to-digital encoders. Generally, they are comprised of a mask attached to the moving mechanical element, along with a means to read the information on the mask. Magnetic, optical, and electrical means are used for reading. Where electrical sensing is used, the mask may consist of a conducting pattern on an insulated substrate. The code represented by the mask pattern is read by means of fixed conducting brushes which are in contact with the pattern. Optical sensors use a light source and photodetectors, whereas magnetic sensors employ inductive-pickup coils.

The sensing means has a finite width. Consequently, there may be an ambiguity in the digital output. This would result, for example, where a conducting brush would be in contact with two adjacent portions of the mask pattern at the same time. Ambiguity can be avoided through the use of special codes and ingenious arrangements of the sensing detectors. Gray code, where only one bit in the digital output changes at any given time as the position of the mask is varied, may be used. A V-scan technique also can be used. In the latter, two detectors are used for each track on the mask. With appropriate decoding of the outputs from the pairs of detectors, an unambiguous digital representation will result.

Electronic A/D Converters: Two classes are used: (1) the input quantity, usually a voltage or current, is converted into another form, such as a frequency or a pulse duration. This intermediate quantity then is measured to yield a digital representation of the input signal. (2) the input is compared directly with a known reference signal which can be varied under control of the A/D converter logic. Several subclasses of A/D converters are obtainable within these two broad classes.

The principal techniques most frequently employed in process control and data-acquisition computers are (1) ramp; (2) integrating-ramp; (3) voltage-to-frequency; (4) successive-approximation; and (5) parallel-serial methods. Check alphabetical index for further coverage. With reference to the two broad classifications, the ramp, integrating ramp, and voltage-to-frequency methods involve the conversion of the input signal into an intermediate quantity before measurement. The successive-approximation and parallel-serial methods are direct-comparison methods.

Numerous cost/performance trade-offs are involved in the selection of the most appropriate A/D converter for a given process control or data-acquisition system. The characteristics are summarized as follows:

Ramp and Voltage-to-Frequency Converters—relatively slow (require serial counting); used at speeds less than several thousand samples/ second; usually resolutions of less than 12 bits.

Successive-Approximation Converters—useful up to about 100,000 samples/second at resolution of 16 bits. Useful up to more than 250,000 samples/second at resolution of 8 bits or less.

Parallel-Serial Converters—high-speed uses requiring conversion rates in excess of 100,000 samples/second; resolution of 8 to 14 bits.

Thomas J. Harrison, International Business Machines
Corporation, Boca Raton, Florida.

ANALOGY (Dynamics). Many dynamical systems, especially mechanical and acoustical ones, can be analyzed by analogy to electrical systems. This is true because these various systems possess certain common attributes such as energy, power, frequency (of vibration, rotation, or alternation), and because equations for calculating these and other quantities can be written in terms of variables that are closely analogous. For example, electromotive force in an electrical system is strictly analogous to force in a mechanical rectilineal system, to torque in a mechanical rotational system, and to pressure in an acoustical system. Again, current in an electrical system is closely analogous to linear velocity in a mechanical rectilineal system, to angular velocity in a mechanical rotational system, and to volume current in an acoustical system.

To show the applications of these analogies the following relationships are given for the four systems cited. Note that a dot above a quantity denotes its first derivative with respect to time.

The kinetic energy T_{KE} stored in the magnetic field of the electrical circuit is

$$T_{KE} = \tfrac{1}{2}Li^2$$

where L = inductance, in abhenries, and i = current through the inductance L, in abamperes.

The kinetic energy T_{KM} stored in the mass of the mechanical rectilineal system is

$$T_{KM} = \tfrac{1}{2}m\dot{x}^2$$

where m = mass, in grams, and $\dot{x}$ = velocity of the mass m, in centimeters per second.

The kinetic energy T_{KR} stored in the moment of inertia of the mechanical rotational system is

$$T_{KR} = \tfrac{1}{2}I\dot{\phi}^2\theta$$

where I = moment of inertia, in gram (centimeter)^2, and ϕ = angular velocity of I, in radians per second.

The kinetic energy T_{KA} stored in the inertance of the acoustical system is

$$T_{KA} = \tfrac{1}{2}M\dot{X}^2$$

where $M = m/S^2$, the inertance, in grams per (centimeter)^4
m = mass of air in the opening, in grams
S = cross-sectional area of the opening, in square centimeters
$X = S\dot{x}$ = volume current, in cubic centimeters per second
$\dot{x}$ = velocity of the air particles in the opening, in centimeters per second

Similar relations hold for potential energy. Thus the potential energy V_{PE} stored in the electrical capacitance of the electrical circuit is

$$V_{PE} = \frac{1}{2}\frac{q^2}{C_E}$$

where C_E = capacitance, in abfarads, and q = charge on the capacitance, in abcoulombs.

The potential energy V_{PM} stored in the compliance or spring of the mechanical rectilineal system is

$$V_{PM} = \frac{1}{2}\frac{x^2}{C_M}$$

where $C_M = 1/s$ = compliance of the spring, in centimeters per dyne
s = stiffness of the spring, in dynes per centimeter (1 newton = 10^5 dynes)
x = displacement, in centimeters

The potential energy V_{PR} stored in the rotational compliance or spring of the mechanical rotational system is

$$V_{PR} = \frac{1}{2}\frac{\phi^2}{C_R}$$

where C_R = rotational compliance of the spring, in radians per dyne per centimeter, and ϕ = angular displacement, in radians.

The potential energy V_{PA} stored in the acoustical capacitance of the acoustical system is

$$V_{PA} = \frac{1}{2}\frac{X^2}{C_A}$$

where X = volume displacement, in cubic centimeters
$C_A = V/\rho c^2$ = acoustical capacitance, in (centimeters)^5 per dyne
V = volume of the cavity, in cubic centimeters
ρ = density of air, in grams per cubic centimeter
c = velocity of sound, in centimeters, per second

Moreover, the total energy in any of these systems can be found simply by adding the kinetic and the potential energies. These relationships are particularly useful in the facility with which relationships like Kirchhoff's law, and its mechanical and acoustical analog, D'Alembert's principle, can be used in the solution of mechanical or acoustical networks, that is, in various series of springs, gears and acoustical elements such as are found in complex trains of equipment as, for example, reduction gearing, loudspeakers, microphones, phonograph pickups, automobile suspensions, wave filters, and other electromechanical and electroacoustic transducers. Even the electrical reciprocity and superposition theorems have their analogs for mechanical and acoustical systems.

The construction of electrical analogies is frequently used in the analysis of process and process control performance. The field of fluidics has brought forth some very interesting and helpful applications of the analog technique.

ANALOGY (Physiology). The relationship between body parts having different embryonic and phylogenetic origin, but with the same function. As an example, the wings of insects and the wings of birds are analogous and illustrate the principle of analogy. These wings are used for the same function, flying, but they arise in an entirely different manner in the embryonic development. See **Homology.**

ANALYSIS (Chemical). Analytical chemistry is that branch of chemistry which is concerned with the detection and identification of the atoms, ions, or radicals (groups of atoms which react as a unit) of which a substance is composed, the compounds which they form, and the proportions of these compounds which are present in a given substance. The work of the analyst begins with sampling, since analyses are performed upon small quantities of material. The validity of the result depends upon the procurement of a sample that is representative of the bulk of material in question (which may be as large as a carload or tankload).

The Revolution in Analytical Chemistry—A Perspective. During the last few decades, not many branches of any science have undergone so much change in the equipment and procedures used as has the field of *chemical analysis*. This revolution also has impacted on how the principles of chemistry are taught today. The revolution in analysis also has had wide influence on technology in general because of the far greater accuracy with which chemical determinations can be made. Just a few years ago, analyses that would yield reliable data in the range of a few parts per million (*ppm*) were considered excellent. The accurate reporting of parts per billion (*ppb*) was achieved in the 1960s. With modern analytical instrumentation available in the 1990s, a part per trillion ($\frac{1}{10}$) sensitivity is achieved for some routine analyses, as, for example, in determinations of the dangerous pollutant, *dioxin*. A special tandem-accelerator mass spectrometer now can detect three atoms of ^{14}C in the presence of 10^{16} atoms of ^{12}C in a radiocarbon age dating procedure. (It is interesting to note that a pinhead would occupy a part per trillion of the area of a road from New York to California; 10^{12} molecules of molecular weight 600 weigh only 10^{-9} gram.)

While the impact of vastly improved chemical analysis has been felt by essentially all phases of science, dramatically more precise data have been of notable significance in the area of pollutants and pharmaceuticals. The effects of minute impurities, beyond detection just a few years ago, now can be determined and become the basis for pollution and drug legislation, litigation, etc. In some cases, unfortunately, the long-term effects of impurities in substances and in the environment remains a pseudoscience of statistics. Consequently instrumentally yielded analytical chemical information requires caution and prudence in its application to decision making at a policy level.

Traditional Analytical Chemistry. In the interest of putting modern analytical chemistry in perspective, it is in order to review that long time period (essentially prior to the 1940s) when the subject was divided into two readily understood areas:

1. *Qualitative chemical analysis*, in which one is concerned simply with the identification of the constituents of a compound or components of a mixture, sometimes accompanied by observations (rough estimates) of whether certain ingredients may be present in major or trace proportions.
2. *Quantitative chemical analysis*, in which one is concerned with the amounts (to varying degrees of precision) of all or frequently of only of some specific ingredients of a mixture or compound. Classically, quantitative chemical analysis is divided into (a) *gravimetric analysis* wherein weight of sample, precipitates, etc., is the underlying basis of calculation, and (b) *volumetric analysis* (titrimetric analysis) wherein solutions of known concentration are reacted in some fashion with the sample to determine the concentration of the unknown. Obviously, the figures from either gravimetric or volumetric determinations are convertible and the two methodologies frequently are combined in a multistep analytical procedure.

Classical laboratory, manual methods conducted on a macroscale where sample quantities are in the range of grams and several milliliters. These are the techniques that developed from the earliest investigations of chemistry and which remain effective for teaching the fundamentals of analysis. However, these methods continue to be widely used in industry and research, particularly where there is a large variety of analytical work to be performed. The equipment, essentially comprised of analytical balances and laboratory glassware, tends to be of a universal nature and particularly where budgets for apparatus are limited, the relative modest cost of such equipment is attractive.

A Gradual Break from Tradition. One of the first breaks from traditional analytical chemistry was the addition of *microchemical methods*. These methods essentially extended macro-scale techniques so that they could be applied for determinations involving very small (milligram) quantities of samples. These methods required fully new approaches or extensive modifications of macro-scale equipment. Consequently, the apparatus usually was sophisticated, relatively costly, and required, greater manipulative skills. Nevertheless, microchemical methods opened up entirely new areas of research, making possible the determination of composition where the availability of samples, as in many areas of biochemistry, was confined to very small quantities.

A second major break was the introduction of *semi-automated analytical apparatus* which introduced an interim step between (a) macroscale and microchemical analysis techniques on the one hand and (b) fully instrumented and automated analytical methods on the other hand. Significant design changes in chemical balances that greatly increased the speed of weighing samples and reagents and automatic and self-refilling burettes are examples of ways in which an analytical procedure could be "tooled" to conserve technician power, reduce drudgery, and often contribute to more reliable and precise results.

Another major break was the introduction of *process analyzers*, which moved the chemical control laboratory from a central location in a materials manufacturing plant to the use of chemical analyzers *on-line*. With this concept, quality control no longer depended upon grab samples, analyzed periodically and thus always behind (time lag) conditions actually occurring in the process itself at any given instant. Many analytical instruments today, at least in principle, are applicable to on-line installation. While thousands of on-line analyzers are in place, usage throughout the processing and manufacturing industries is far from universal. Difficulties in designing and protecting sensitive instrumentation from the very rugged environments encountered on-line continue. Thus, chemical composition is commonly inferred from other related measurements, such as temperature, pressure, and careful chemical analysis of raw materials at the input side and similar analyses of products on the output side. Numerous techniques used for on-line instrumentation are described in this encyclopedia.

Energy-Matter Interactions in Analytical Instrumentation. Modern chemical analyzers, ranging from research and laboratory applications to process control, developed in a rather chaotic manner over several decades. There indeed was no master plan and, in fact, it was not until the late 1950s that a concerted attempt was made to classify analytical instruments in a scientific way. The accompanying table is an updated, but abridged version of a summary prepared and first published in 1957.[1] The thrust of the summary is directed toward industrial instrumentation, although it embraces the principal laboratory instruments as well.

---

[1] Albright, C. M., Jr.: "Chemical Composition," in *Process Instruments and Controls Handbook*, D. M. Considine, Editor, McGraw-Hill, New York, 1957. The book is now in its 4th edition (1993).

INTERACTIONS BETWEEN ENERGY AND MATTER UTILIZED IN ANALYTICAL INSTRUMENTATION

GROUP I—INTERACTIONS WITH ELECTROMAGNETIC RADIATION

Measurement of the quantity and quality of electromagnetic radiation emitted, reflected, transmitted, or diffracted by the sample.

Electromagnetic radiation varies in energy with radiation frequency, that of the highest frequency or shortest wavelength having the highest energy and penetration into matter. Radiation of the shortest wavelengths (gamma rays) interacts with atomic nuclei; x-rays with the inner shell electrons; visible and ultraviolet light with valence electrons and strong interatomic bonds; and infrared radiation and microwaves with the weaker interatomic bonds and with molecular vibrations and rotation. Most of these interactions are structurally related and unique. They may be used to detect and measure the elemental or molecular composition of gas, liquid, and solid substances within the limitations of available equipment.

Emitted Ratiation
- Thermally Excited: Optical emission spectrochemical analysis
  - Flame photometry
- Electromagnetically Excited: Fluorescence
  - Raman spectrophotometry
  - Induced radioactivity
  - X-ray fluorescence

Transmitted and Reflected Radiation
- X-ray analysis
- Ultraviolet spectrophotometry
- Ultraviolet absorption analysis
- Conventional photometry—transmission colorimetry
- Colorimetry
- Light scattering techniques
- Optical rotation—polarimetry
- Refractive index
- Infrared spectrophotometry
- Infrared process analyzers
- Microwave spectroscopy
- Gamma ray spectroscopy
- Nuclear quadrupole moment

GROUP II—INTERACTION WITH CHEMICALS

Measurement of the results of reaction with other chemicals in terms of amount of sample or reactant consumed, product formed, or thermal energy liberated, or determination of equilibrium attained.

The selectivity inherent in the chemical affinity of one element or compound for another, together with their known stoichiometric and thermodynamic behavior, permits positive identification and analysis under many circumstances. In a somewhat opposite sense, the apparent dissociation of substances at equilibrium in chemical solution gives rise to electrically measureable valence potentials, called oxidation-reduction potentials, whose magnitude is indicative of the concentration and composition of the substance. While individually all the above effects are unique for each element or compound, many are readily masked by the presence of more reactive substances so they can be applied only to systems of known composition limits.

Consumption of Sample or Reactant
- Orsat analyzers
- Automatic titrators

Measurement of Reaction Products
- Impregnated paper-tape devices
- Photometric reaction product analyzers

Thermal Energy Liberation
- Combustion-type analyzers
- Total combustibles analyzers (hydrocarbons and carbon monoxide analyzers)

Equilibrium Solution Potentials
- Redox potentiometry
- pH (hydrogen ion concentration)
- Metal ion equilibria

GROUP III—REACTION TO ELECTRIC AND MAGNETIC FIELDS

Measurement of the current, voltage, or flux changes produced in energized electric and magnetic circuits containing the sample.

The production of net electric charge on atoms or molecules by bombardment with ionizing particles or radiation or by electrolysis or dissociation in solution or the induction of dipoles by strong fields establishes measurable relationships between these ionized or polarized substances and electric and magnetic energy. Ionized gases and vapors can be accelerated by applying electric fields, focused or deflected in magnetic fields, and collected and measured as an electric current in mass spectroscopy. Ions in solution can be transported, and deposited if desired, under the influence of various applied potentials for coulometric or polarographic analysis and for electrical conductivity measurements. Inherent and induced magnetic properties give rise to specialized techniques, such as oxygen analysis based on its paramagnetic properties and nuclear magnetic resonance, which is exceedingly precise and selective for determination of the compounds of many elements.

Mass Spectroscopy
- Quadrupole mass spectrometry

Electrochemical
- Reaction product analyzers

Electrical Properties
- Electrical conductivity/electrical resistivity
- Dielectric constant and loss factor
- Oscillometry
- Gaseous conduction

Magnetic Properties
- Paramagnetism
- Nuclear magnetic resonance
- Electron paramagnetic resonance

GROUP IV—INTERACTION WITH THERMAL OR MECHANICAL ENERGY

Measurement of the results of applying thermal or mechanical energy to a sample in terms of energy transmission, work-done, or changes in physical state.

The thermodynamic relationship involving the physical state and thermal energy content of any substance permits analysis and identification of mixtures of solids, liquids, and gases to be based on the determination of freezing or boiling points and on the quantitative measurement of physically separated fractions. Useful information can often be derived from thermal conductivity and viscosity measurements, involving the transmission of thermal and mechanical energy, respectively.

Effects of Thermal Energy
- Thermal conductivity
- Melting and boiling point determinations
- Ice point-humidity instrumentation, among others
- Dew point-humidity instrumentation, among others
- Vapor pressure
- Fractionation
- Chromatography
- Thermal expansion

Effects of Mechanical Energy or Forces
- Viscosity
- Sound velocity
- Density and specific gravity

Chemical-composition variables are measured by observing the interactions between matter and energy. That such measurements are possible stems from the fundamental that all known matter is comprised of complex, but systematic arrangements of particles which have mass and electric charge. Thus, there are neutrons which have mass but no charge; protons which have essentially the same mass as neutrons with a unit positive charge; and electrons which have a negligible mass with a unit negative charge. The neutrons and protons comprise the nuclei of atoms. Each nucleus ordinarily is provided with sufficient orbital electrons, in what is often visualized as a progressive shell-like arrangement of different energy levels, to neutralize the net positive charge on the nucleus. The total number of protons plus neutrons determines the atomic weight. The number of protons which, in turn, fixes the number of electrons, determines the chemical properties and the physical properties, except mass, of the resulting atom.

The chemical combinations of atoms into molecules involve only the electrons and their energy states. Chemical reactions involving both structure and composition generally occur by loss, gain, or sharing of electrons among the atoms. Thus, every configuration of atoms in a molecule, crystal, solid, liquid, or gas may be represented by a specific system of electron energy states. Also, the particular physical state of the molecules, as resulting from their mutual arrangement, also is reflected upon these energy states. Fortunately, these energy states, characteristic of the composition of any particular substance, can be inferred by observing the consequences of interaction between the substance and an external source of energy.

External energy sources used in analytical instrumentation include:

1. electromagnetic radiation
2. electric or magnetic fields
3. chemical affinity or reactivity
4. thermal energy
5. mechanical energy

The interaction of electromagnetic radiation with matter yields fundamental information as the result of the fact that photons of electromagnetic radiation are emitted or absorbed whenever changes take place in the quantized energy states occupied by the electrons associated with atoms and molecules. X-rays (photons or electromagnetic wave packets with relatively high energy) penetrate deeply into electron orbits of an atom and provide, upon absorption, the large quantity of energy required to excite one of the innermost electrons. Thus, the pattern of x-ray excitation or absorption is relative to the identity of those atoms whose orbital electrons are excited, ideally suiting x-ray techniques for determining atoms and elements in dense samples. But, because of the penetrating power of x-rays, they are not suited to the excitation of low-energy states which correspond to outer-shell or valence electrons; or of the interatomic bonds which involve vibration or rotation.

In contrast, the relatively longer wavelengths of infrared radiation (photons having relatively low energy) correspond to the energy transformations involved in the vibration of atoms in a molecule as resulting from stretching or twisting of the interatomic bonds. Thus, because the penetrating power of electromagnetic radiation varies over the total spectrum, an instrumental irradiation technique can be developed for almost any analytical instrumentation requirement.

The interaction of matter with electric or magnetic fields is widely applied for determining chemical composition. The mass spectrometer, for example, which uses a combination of electric and magnetic fields to sort out constituent ions in a sample, takes full advantage of this interaction. A simple electric-conductivity apparatus determines ions in solution as the result of applying an electric potential difference across an electrolyte.

The use of chemical reactions in analytical instrumentation essentially extends the fundamental techniques of older laboratory analytical methods.

Numerous analytical instrumentation techniques involve interactions between mechanical and thermal energy with matter. All of these interactions are summarized in the accompanying table.

Targets of Analytical Instrumentation. Some authorities feel that less emphasis and even abandonment by some educational institutions of the traditional qualitative-analysis (wet basis) course represents the loss of a great learning experience in the fundamentals of chemistry. Generally, for teaching purposes, the course is limited to inorganic sustances. Practically all of the fundamentals of inorganic chemistry are called upon in the execution of qualitative analysis. Thus, in addition to serving as an effective analytical procedure, the method is an effective teacher.

The first important step is that of putting the sample (unknown) into solution. For metals and alloys, strong acids, such as HCl, HNO_3, or aqua regia may be used. If the material is not fully dissolved by these acids, it should be fused, either with sodium carbonate (alkaline fusion) or potassium acid sulfate (acid fusion). Care should be exercised to make certain that no portion of the unknown is volatilized and thus lost during these procedures.

The next step is the detection of the cations of the metals. For this purpose, the solution should be treated with HNO_3, by evaporation and redissolving if necessary, to remove other acid radicals, so that nitrate is the only anion present in the solution. Then, a systematic procedure is followed for separation of groups of the cations. Such schemes of separation have been devised for all the metals found in nature. A shortened plan, which applies to 24 of the commonly occurring metals and ammonium, has been known and practiced for many years. This plan consists of the separation of the 24 metals into five groups. Further details of the procedure are given in the 6th Edition of this encyclopedia.

Criteria for Selecting Appropriate Analytical Method. These include (1) sensitivity, (2) specificity, (3) speed, (4) sampling methodology required, (5) simplicity (translated into terms of expertise needed), and, of course, cost as traded off against the other criteria.

Sensitivity, as previously mentioned, has improved almost astoundingly over the past few years. Of course, great sensitivity is not always needed, with the ppm and ppb levels well serving many industrial and laboratory requirements. Most frequently, sensitivity is closely related to cost. Sensitivity is also closely related to accuracy, i.e., sensitivity of a reliable, repeatable nature. See also separate entries on **Accuracy; Repeatability; Reproducibility;** and **Sensitivity,** and check alphabetical index.

Sensitivity depends on how well the target of measurement can be transduced into some reliable signal from which the instrument can create a display and/or record. The absorption or emission of photons is the basis of many spectroscopic analytical methods, such as x-ray, ultraviolet, infrared, nuclear magnetic resonance, Raman, Mossbauer, etc., as well as of charged particles, which serve as the basis for electron and mass spectrometry and of the electrochemical, flame ionization, etc. used in chromatography. Through the use of intense energy sources, such as lasers, synchrotron radiation, and plasmas, the efficiency of converting (transducing) the objective parameter of analysis (*analyte*) is greatly improved. McLafferty reports that efficiencies approaching 100% have been experienced for resolution-enhanced multiphoton ionization of atomic and molecular species. Multiplier detectors can respond to the arrival of a single photon or ion. Such methods can detect, for example, a single cesium atom or naphthalene molecule.

Specificity is an everpresent criterion because there are indeed few analytical techniques that detect single species without careful tuning. Frequently, filtering techniques must be used as a means of narrowing the range of detection. See **Infrared Radiation;** and **Ultraviolet Spectrometers.**

Speed. The rapidity with which an analysis can be performed and utilized (including interpretation, whether manual or automatic) is particularly important in industrial chemical analysis. In a laboratory setting, this may not be quite so urgent, but even then time is a major criterion where, in most cases, special personnel are held up in other activities, awaiting the results of an analysis. Frequently higher cost can be justified on the basis of less time and lower personnel costs per analysis made.

One of the principal contributions of electronic data processing over the past several years in terms of chemical analysis is the savings of manual effort in interpreting analytical data. Special techniques, such as Fourier transform, have increased speed (as well as sensitivity) by orders of magnitude in connection with infrared, nuclear magnetic resonance, and mass spectroscopy. Of course, for on-line process analyses, essentially instantaneous interpretation is required to provide the proper error signal that is used to position the final control element (valve, feeder, damper, etc.).

Sampling for analysis is sometimes considered a secondary criterion in the selection of an analysis system. In the laboratory or on the process, sampling often is the practical key to success. The entire result can depend upon obtaining a *truly representative* sample and the sampling methodology used varies widely with the materials to be analyzed. Sampling of solids, such as coal, metals, etc. differs markedly from sampling for fluids. Particularly in process analyses, where the environment varies and contrasts markedly with the usual laboratory conditions, a gas or liquid will require filtering and temperature and pressure conditioning—so that the composition detector will consistently be exposed to the material in question under the same physical conditions.

Additional Reading

Anderson, D. J.: "Analysis in Clinical Chemistry," *Analytical Chemistry*, 165R (June 15, 1991).

Anderson, D. G.: "Analysis of Coatings," *Analytical Chemistry*, 87R (June 15, 1992).

Brettell, T. A., and R. Saferstein: "Analysis in Forensic Science," *Analytical Chemistry*, 148R (June 15, 1991).

Cherry, R. H.: "Thermal Conductivity Gas Analyzers," in *Process/Industrial Instruments and Controls Handbook*, D. M. Considine, Editor, McGraw-Hill, New York, 1993.

Clement, R. E., Langhorst, M. L., and G. A. Eiceman: "Environmental Analysis," *Analytical Chemistry*, 270R (June 15, 1991).

Converse, J. G.: "Sampling for On-Line Analyzers," in *Process/Industrial Instruments & Controls Handbook*, D. M. Considine, Editor, 4th Edition, McGraw-Hill, New York, 1993.

Converse, J. G.: "Process Chromatography," in *Process/Industrial Instruments & Controls Handbook*, D. M. Considine, Editor, 4th Edition, McGraw-Hill, 1993.

Dulski, T. R.: "Analysis of Steel and Related Materials," *Analytical* Chemistry, 65R (June 15, 1991).

Foucault, A. P.: "Countercurrent Chromatography," *Analytical Chemistry*, 569A (May 15, 1991).

Fox, D. L.: "Analysis for Air Pollution," *Analytical Chemistry*, 29R (June 15, 1991).

Gilpin, R. K., and L. A. Pachla: "Analysis of Pharmaceuticals and Related Drugs," *Analytical Chemistry*, 130R (June 15, 1991).

Glajch, J. L., and L. R. Snyer, Editors: "Computer-Assisted Method Development for High-Performance Liquid Chromatography," Elsevier Science, New York, 1990.

Graham, J. A.: "Monitoring Groundwater and Well Water for Crop Protection Chemicals," *Analytical Chemistry*, 613A (June 1, 1991).

Harman, J. N., and D. M. Gray: "pH and Redox Potential Measurements," in *Process/Industrial Instruments & Controls Handbook*, D. M. Considine, Editor, 4th Edition, McGraw-Hill, New York, 1993.

Jackson, L. L.: "Analysis of Geological and Inorganic Materials," *Analytical Chemistry*, 33R (June 15, 1991).

Kohlmann, F.: "Electrical Conductivity Measurements," in *Process/Industrial Instruments & Controls Handbook*, D. M. Considine, Editor, 4th Edition, McGraw-Hill, New York, 1993.

Lex, D.: "Turbidity Measurement," in *Process/Industrial Instruments & Controls Handbook*, D. M. Considine, Editor, 4th Edition, McGraw-Hill, New York, 1993.

MacCarthy, P., et al.: "Water Analysis," *Analytical Chemistry*, 301R (June 15, 1991).

MacLeod, S. K.: "Moisture Determination Using Karl Fischer Titrations," *Analytical Chemistry*, 557A (May 15, 1991).

McManus, T. R.: "Analysis of Petroleum and Coal," *Analytical Chemistry*, 48R (June 15, 1991).

Nadkarni, R. A.: "The Quest for Quality in the Laboratory," *Analytical Chemistry*, 675A (July 1, 1991).

Newman, A. R.: "Electronic Noses," *Analytical Chemistry*, 588A (May 15, 1991).

Newman, A. R.: "Portable Analytical Instruments," *Analytical Chemistry*, 641A (June 1, 1991).

Ondov, J. M., and W. R. Kelly: "Tracing Aerosol Pollutants with Rare Earth Isotopes," *Analytical Chemistry*, 691A (July 1, 1991).

Ray, M. A.: "Surface Characterization (Chemical Substances and Products)," *Analytical Chemistry*, 99R (June 15, 1991).

Saltzman, R. S.: "Gas and Process Analyzers," in *Process/Industrial Instruments & Controls Handbook*, D. M. Considine, Editor, 4th Edition, McGraw-Hill, New York, 1993.

Sherma, J.: "Analysis of Pesticides," *Analytical Chemistry*, 118R (June 15, 1991).

Smith, C. G., et al.: "Analysis of Synthetic Polymers," *Analytical Chemistry*, 11R (June 15, 1991).

Tipping, F. T.: "Oxygen Determination," in *Process/Industrial Instruments and Controls Handbook*, D. M. Considine, Editor, 4th Edition, McGraw-Hill, New York, 1993.

Wade, A. P., et al.: "An Analytical Perspective on Acoustic Emission," *Analytical Chemistry*, 497A (May 1, 1991).

Yazbak, G.: "Refractomers," in *Process/Industrial Instruments & Controls Handbook*, D. M. Considine, Editor, 4th Edition, McGraw-Hill, New York, 1993.

ANALYSIS OF COVARIANCE. A generalization of *Analysis of Variance* to the case where more than one variable is observed on each member of the sample. Suppose, for example, there are two variables x and y such as the scores on a test of a group of students before undergoing a course of instruction (x) and after the course is completed (y). The final performance y will be influenced both by the course and the knowledge of the student at the outset, represented by x. To disentangle these effects y is regressed on x and the residual $y - bx$ computed, b being the regression coefficient. This residual should, in suitable circumstances, represent the effect of the course regardless of initial knowledge and the set of residuals can be subjected to variance analysis if the students are classified in any way. More generally if x is a variable unaffected by classification or treatment in an experiment its effect on y can be extracted from y and the residuals analyzed in the ordinary way; and so for several variables of type x, which can also be abstracted from y by the use of a regression equation.

Sir Maurice Kendall, International Statistical Institute, London.

ANALYSIS OF VARIANCE. In statistics, a technique for segregating the causes of variability affecting a set of observations. Consider a simple case in which a number of observations are taken on members falling into different classes (for example the yields of a number of plots of wheat, groups of which are subjected to different fertilizer treatments). The problem is whether the yields differ from group to group or, on the other hand, differ only as random variations from a homogeneous population. The matter is decided by comparing the sum of squares of mean yields of groups about the overall mean of all plots with the aggregated sum of squares of observations within groups about their respective group means. On the assumption that the variation is normal (Gaussian) an exact test of significance can be applied to decide whether the difference is great enough to justify the conclusion that group differences are real.

The method can be generalized to much more elaborate situations where the classifications are more complex, especially in experimental designs which are carefully balanced so that analyses of variance are easy to apply.

The technique is frequently referred to as ANOVA.

Sir Maurice Kendall, International Statistical Institute, London.

ANALYSIS (Organic Chemical). Various techniques are used in the chemical analysis of organic substances both in microanalysis and macro laboratory procedures. As contrasted with the determination of total carbon content or the amounts of other specific chemical elements, the representative analytical techniques described here are directed toward the determination of presence and amount of various functional groups (radicals). These groups also are described elsewhere in this volume and, in several instances, additional analytical procedures are related.

(1) The *carboxyl group* is determined by titration with standard sodium hydroxide solution, using phenolphthalein as the indicator, by the reaction

$$\mathrm{RCOOH + NaOH \rightarrow RCOONa + H_2O}$$

(2) The *hydroxyl group* is determined by reaction with acetic anhydride on heating in a sealed tube by the reaction

$$\mathrm{ROH + (CH_3CO)_2O \rightarrow ROOCCH_3 + CH_3COOH}$$

The amount of hydroxyl group present is found by titrating the resulting acetic acid (CH_3COOH) with standard sodium hydroxide, as in (1).

(3) The *acyl group* (—COOR) in esters and amides is determined by hydrolysis in alcoholic sodium hydroxide solution, followed by ion exchange with an acidic resin. The carboxylic acid formed is then titrated with standard sodium hydroxide, as in (1). The reactions are

$$\underset{\text{Ester}}{\mathrm{RCOOR'}} + \mathrm{NaOH \rightarrow RCOONa + R'OH}$$

or

$$\underset{\text{Amide}}{\mathrm{RCONHR'}} + \mathrm{NaOH \rightarrow RCOONa + R'NH_2 + H_2O}$$

$$\mathrm{RCOONa + Resin\text{-}SO_3H \rightarrow RCOOH + Resin\text{-}SO_3Na}$$

(4) The *carbonyl group* is determined by a reaction with 2,4-dinitrophenyl hydrazine which precipitates the 2,4-dinitrophenyl hydrazone of the aldehyde or ketone, which is then filtered off, dried, and weighed. The reaction is

$$\mathrm{RR'CO + H_2NNHC_6H_3(NO_2)_2 \rightarrow RR'C{:}NNHC_6H_3(NO_2)_2 + H_2O}$$

(5) The *peroxy group* is determined by treatment with sodium iodide. The liberated iodine is then titrated with standard sodium thiosulfate solution. The reaction is

$$\mathrm{RCOO_2OCR + 2NaI \rightarrow I_2 + 2R'COONa}$$

(6) The primary *amino group* is determined by treatment with nitrous acid and measurement of the nitrogen (gas) produced by the reaction

$$\mathrm{RNH_2 + HNO_2 \rightarrow N_2 + ROH + H_2O}$$

(7) The aromatic *nitro group* is determined by its reduction with excess titanium(III) chloride. After the reaction, the unused titanium(III) ions (Ti^{3+}) are determined by titration with iron(III) sulfate or iron alum solution:

$$RNO_2 + 6TiCl_3 + 6HCl \rightarrow RNH_2 + 6TiCl_4 + 2H_2O$$

$$Ti^{3+} + Fe^{3+} \rightarrow Ti^{4+} + Fe^{2+}$$

(8) The *hydrazino group* is determined by oxidation with copper(II) sulfate solution, and measurement of the nitrogen (gas) formed. The reaction is

$$RNHNH_2 + 4CuSO_4 + H_2O \rightarrow N_2 + ROH + 2Cu_2SO_4 + H_2SO_4$$

(9) The *sulfhydryl group* is determined by reaction with iodine, which is produced in the vessel from potassium iodide, added in excess to the solution, and potassium iodate, added from a buret until the completion of the reaction is shown by the permanent appearance of the blue color of starch-iodine.

$$2RSH + I_2 \rightarrow RSSR + 2HI$$

(10) *Unsaturated groups* are determined by addition of bromine, by the reaction

$$R_2C{=}CR_2' + Br_2 \rightarrow R_2CBr{-}CBrR_2'$$

The term *functional group analysis* sometimes is used to describe the foregoing kinds of analyses.

Ultimate Analysis. This term, generally limited to organic chemical analysis, denotes the determination of the proportion of each element in a given substance. The primary determination is that of carbon and hydrogen, which is conducted by mixing the sample with copper(II) oxide and heating it in a stream of oxygen to a temperature of 700 to 800°C. The carbon is converted to carbon dioxide and the hydrogen to water. These products are then absorbed by suitable reagents. For example, magnesium perchlorate dehydrate may be used to absorb water and sodium hydroxide to absorb carbon dioxide. Although the fundamental procedure is simple, a rather elaborate train of apparatus, involving both temperature and flow control, is required. The traditional procedure for determining nitrogen is the Kjeldahl method.

In the Unterzaucher method for determining oxygen in organic substances, the sample is heated to a high temperature (approximately 1120°C) in an atmosphere of nitrogen. Under these conditions, the oxygen present combines with part of the carbon content to form carbon dioxide and with part of the hydrogen content to form water. The gases then are passed over hot carbon (1150°C), whereupon both the carbon dioxide and water are converted to carbon monoxide. The latter gas upon leaving the furnace is passed over iodine pentoxide I_2O_5 at about 110°C to form iodine by the reaction: $5CO + I_2O_5 \rightarrow I_2 + 5CO_2$. The freed iodine is titrated with a standard sodium thiosulfate solution.

ANALYTICAL BALANCE. See **Weighing.**

ANALYTICAL GEOMETRY. See **Geometry.**

ANALYTIC CONTINUATION. Calculation of an analytic function over some domain, from precise definition of the function over a smaller domain.

Suppose $f_1(z)$ is analytic in D_1 and $f_2(z)$ in D_2 and that D has a region in common with both D_1 and D_2. Further suppose that $f_1(z) = f_2(z)$ in D, then if $f(z)$ can be defined so that $f(z) = f_1(z)$ in D_1 and $f(z) = f_2(z)$ in D_2 the analytic continuation of either $f_1(z)$ or $f_2(z)$ in the domain ($D_1 + D_2$) is $f(z)$. As a simple example consider the series ($1/a + z/a^2 + z^2/a^3 + \cdots$) which represents the function $1/(a - z)$ only within C_1, a circle of radius $|a|$. Another power series of the type $[1/(a - b) + (z - b)/(a - b)^2 + (z - b)^2/(a - b)^3 + \cdots]$, however, represents the same function outside C_1 if b/a is not real and positive. This series converges at points inside another circle which has regions in common with C_1. See also **Taylor Series.**

ANALYTIC FUNCTION. A function $f(z)$ of the complex variable $z = x + iy$ is analytic at a point on the z-plane if the function and its first derivative are finite and single-valued there. If this property applies to all points within a given region of the complex plane, $f(z)$ is an analytic function throughout the region. Any point at which the derivative fails to exist is a singularity or a singular point of the function. According to the Liouville theorem, if $f(z)$ has no singularity for z finite or infinite it is a constant.

Equivalent definitions of an analytic function are: (1) it must satisfy the Cauchy-Riemann equations and Laplace's equation; (2) it is analytic only if it may be represented by a convergent power series in some neighborhood of the given point.

Other words often used in place of analytic, and essentially equivalent, are holomorphic, meromorphic, monogenic, uniform, regular.

An analytic function of a real variable may be defined in a similar way. See also **Cauchy Theorem.**

ANALYZER (Optics). A term applied to the Nicol prism (or other device which passes only plane polarized light) which is placed in the eyepiece of a polariscope or similar instrument.

ANALYZER (Reaction-Product). Chemical composition may be determined by the measurement of a reaction product—in an automatic fashion utilizing the basic principles of conventional qualitative and quantitative chemical analysis. Two steps usually are involved in this type of instrumental analysis: (1) the formation of a target chemical reaction, and (2) the determination of one or more of the reaction products.

Determination of a constituent in a process stream or sample by measurement of a reaction product can be represented by: $C + R \rightarrow P$, where C = constituent to be determined; R = reactant; and P = reaction product to be measured. If reactant R already is present in the sample, it is only required to expose the sample to suitable reaction conditions to form P. Under normal instrument operating conditions, the reaction of C and R may be spontaneous. In other instances, suitable conditions may have to be established either (1) to promote the desired reaction (for example, setting the proper temperature and pressure, or using a catalyst) or (2) to assure a suitable reaction rate.

Frequently, it is possible to measure the reaction product as it forms in the reaction zone. In some instances, the products and sample residue must be removed from the reaction zone before a measurement can be made. Also, the reaction product may be measured directly; or its presence may have to be inferred from a secondary reaction. In one example, carbon monoxide in air or oxygen may be determined by combustion to carbon dioxide. The latter may be measured directly as by thermal-conductivity methods; or inferentially by absorbing the carbon dioxide in a solution and then measuring the change of that solution by electrolytic conductance.

ANALYZER (Reagent-Tape). The key to chemical analysis by this method is a tape (paper or fabric) that has been impregnated with a chemical substance that reacts with the unknown to form a reaction product on the tape which has some special characteristic, e.g., color, increased or decreased opacity, change in electrical conductance, or increased or lessened fluorescence. Small pieces of paper treated with lead acetate, for example, have been used manually by chemists for many years to determine the presence of hydrogen sulfide in a solution or in the atmosphere. This basic concept forms the foundation for a number of sophisticated instruments that may pretreat a sample gas, pass it over a cyclically advanced tape, and, for example, photometrically sense the color of the exposed tape, to establish a relationship between color and gas concentration. Depending upon the type of reaction involved, the tape may be wet or dry and it may be advanced continuously or periodically. Obviously, there are many possible variations within the framework of this general concept.

ANAMNIA. Vertebrates which do not develop an amnion during embryonic life. The group includes the cyclostomes (see **Cyclostomata**), fishes, and amphibians.

ANAMORPHISM. A term proposed by Van Hise in 1904 to designate the deep-seated constructive processes of metamorphism by which new complex (metamorphic) minerals are formed from the pre-existing simpler minerals, as contrasted with the surface alteration of rocks due to weathering and cementation, termed katamorphism.

ANAPHYLAXIS. State of supersensitivity which may develop after a first injection of a foreign protein, such as a therapeutic or prophylactic serum. See also **Alkaloids.**

ANA-POSITION. The position of two substituent groups on atoms diagonally opposite, in α-positions on symmetrical fused rings, as the 1,5 or the 4,8 positions (which are identical) of the naphthalene ring.

ANASTIGMAT. A compound lens combination corrected so that both astigmatism and the curvature of the field are largely eliminated over a considerable area in the image plane.

ANATASE. The mineral anatase, TiO_2 crystallizing in the tetragonal system is a relatively uncommon mineral. It occurs as a trimorphous form of TiO_2 with rutile and brookite. Rutile and anatase have tetragonal crystallization; brookite, orthorhombic. It was originally named octahedrite from its pseudo-octahedral, acute pyramidal crystal habit. Hardness, 5.5–6; sp. gr. 3.82–3.97; brittle with subconchoidal fracture; color, shades of brown, into deep blue to black; also colorless, grayish, and greenish. Transparent to opaque with adamantine luster.

Anatase occurs as an accessory mineral in igneous and metamorphic rocks, gneisses, and schists. Fine crystals have been found in Arkansas in the United States, and in Switzerland.

ANATEXIS. A term proposed by Sederholm in 1907 for the supposed end-process of deep-seated metamorphism resulting in the partial or complete remelting of a specific type of rock in situ.

ANATOMY. A branch of biology dealing with structure, generally considered to be gross structure, but sometimes used to refer to microscopic structure as well. A subdivision of the more inclusive term, morphology, which includes all forms of study of structure. Human anatomy is a study of the various organs of the human body and their relationship to each other as to shape and position.

Classically, anatomy has been divided into a number of subclasses: (1) *gross anatomy* which is a study of macroscopic structure, that is, the structure which can be seen with the unaided eye; (2) *comparative anatomy* which studies the structures of animals in relation to each other, including human structure; (3) *developmental anatomy* which studies both embryonic and later development of body structures; (4) *functional anatomy* which studies the interaction of organs, particularly as they change in shape, size, pressure, temperature, and other important ways; (5) *microscopic anatomy* (histology) which investigates minute structure, particularly of cells in tissues and how cells develop into organs; and (6) *pathological anatomy* which studies diseased structures, that is, the deviation from normal structures and functions. There also are the classical subclasses of *human anatomy, animal anatomy*, and *plant anatomy.*

- Abdominal Cavity
  - Gallbladder
  - Intestines
  - Kidneys
  - Liver
  - Pancreas
  - Pelvic Cavity
    - Bladder
    - Pelvis
    - Rectum
  - Spleen
  - Stomach
- Buccal Cavity
  - Teeth
  - Tongue
- Cranial Cavity
  - Brain
- Nasal Cavity
  - Structures forming the nose
- Orbital Cavities
  - Eyes, eyeball muscles
  - Lacrimal apparatus
  - Optic nerves
- Peritoneal Cavity
- Pleural Cavities
- Spinal Canal
  - Spinal Cord
- Thoracic Cavity
  - Blood and lymph vessels
  - Esophagus
  - Heart
  - Lungs
  - Trachea
  - Thymus gland

Principal cavities of the body, indicating what they contain.

Anatomists have likened the human body (and other vertebrates) to a tube which may be referred to as the *body wall;* this tube enclosing another tube referred to as the *viscera.* The cavity between the tubes may be referred to as the *body cavity* or *celom.* The principal cavities of the human body are outlined in the foregoing list. The major systems of the body are: (1) the *circulatory* or *vascular system* (blood, blood vessels, heart, lymphatic vessels and lymph); (2) the *digestive system* (alimentary canal, pancreas, liver, salivary gland; (3) the *endocrine system* (adrenals, parathyroids, pituitary, thyroid, portions of ovaries and testes, and other glands with ducts); (4) the *excretory system* (bladder, kidneys, ureters, urethra, respiratory systems of the skin); (5) *muscular system;* (6) *nervous system* (brain, ganglia, nerve fibers, spinal cord); (7) *reproductive system* (bulbourethral and prostate glands, penis, seminal vesicles, testes, and urethra in the male; ovaries, uterine tubes, vagina, and vulva in the female); (8) the *respiratory system* (bronchi, larynx, lungs, nose, pharynx); and (9) the *skeletal system* (bones and connective tissue).

Needless to say, the dividing line between anatomy, physiology, and other biological and medical sciences is indistinct and growing less distinct as scientists emphasize the interdisciplinary approach to their work.

ANCHOR RING (or Torus). A surface that has the shape of a doughnut. It can be generated as a surface of revolution by rotating the circle.

$$(y - b)^2 + z^2 = a^2$$

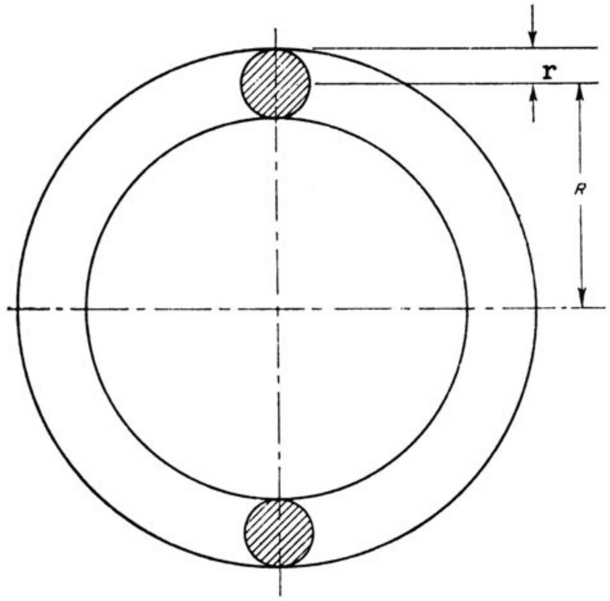
around the Z-axis. Its equation, when rationalized, is of the fourth degree

$$(x^2 + y^2 + z^2 + b^2 - a^2)^2 = 4b^2(x^2 + y^2)$$

With dimensions as shown in the figure, its volume, $V = 2\pi^2Rr^2$ and its surface area, $A = 4\pi^2Rr$.

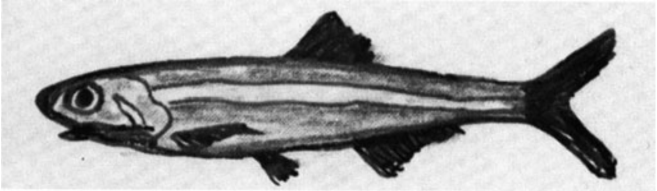
This surface is of interest in topology, where it is said to be of genus 1.

See also **Circle (Geometry);** and **Topology.**



Anchor ring.

ANCHOVY AND ANCHOVETA (*Osteichthyes*). The anchovy family (*Engraulidae*) comprises 15 genera and some 100 species. These fishes are found in the tropical and temperate regions of the northern and southern hemispheres. Distribution is chiefly in the Indian and Pacific Oceans. Anchovies school along the coast and some are also found in fresh water. Because of the great masses in which they occur, the family has large commerical importance to several countries for the



Striped anchovy (*Anchoa hepsetus*).

production of fish meal and oil. One of the main differences between the anchovy and the herring is the prominent protruding upper jaw as indicated in the accompanying illustration.

Anchovies. Seven species in the main genus *Engraulis* have been identified from the Pacific and Atlantic Oceans. Included is the anchovy (*Engraulis encrasicholus*) which reaches about 8 inches (20 centimeters) in length, but is usually from 4.7 to 6.3 inches (12 to 16 centimers) in length. Coloration resembles that of the herring, with silver lateral stripes. Distribution is chiefly in the Mediterranean Sea and Black Sea as well as on the Atlantic coast of southwest Europe and north Africa. In the north, the distribution extends to the English Channel, and to the south along the African west coast from Togo to Dahomey. Anchovies are also found in the Sea of Asov, in the southern North Sea, and in small numbers as far North as Bergen, Norway.

In their chief distribution region, e.g., in the Mediterranean, anchovies migrate little. In spring and summer, they appear in great schools at the surface of the water both in the open sea and along the coast. They spawn in the Mediterranean from April to September. During this time, they are also fished. After spawning, at the commencement of winter, the adults and the subadults from the spring and summer spawn move into depths of 329 to 492 feet (100 to 150 meters). The schools probably break up at this time and the fish remain in some small region at the floor of the sea. This is evident from studies made of their stomach contents.

Anchovies in the bordering regions of their distribution are migratory. They migrate from wintering grounds in the Black Sea in early spring to the Sea of Asov, where they spawn and return to the Black Sea in the fall. In northern regions, anchovies also migrate in great schools to the north and northeast. They move through the Bristol Channel into the Irish Sea and to the west coast of Scotland, where they can be found from May to September. Spawning grounds are presumably located in this area, since specimens found there are mature for spawning.

In spring, the anchovies migrate in great numbers through the English Channel to the North Sea. They move along the French-Belgian-Dutch coast to the East Frisian coast. The spawning grounds were located here in 1929, and considerable fishing activity developed there, particularly in the Zuider Zee. Since that time, particularly after World War II, it was observed in the southeastern North Sea that the number of anchovies decreased and that they also spawned there. Eggs and larvae were found as far as the North Frisian islands. The spawning period was in warmer brackish water from June to August. Climatic changes probably acted as a factor in the spread of anchovies to the north, as was the case with sardines. In the fall, the anchovies migrate through the northern part of the English Channel along the English coast to wintering grounds off the west exit of the Channel.

The number of eggs varies between 13,000 and 20,000; they are laid in open water in groups. After a year, the fish are 3.5 to 4 inches (9 to 10 centimeters) long and spawn for the first time in the following summer at a length of about 4.75 and 5 inches (12 to 13 centimeters). The Sea of Asov anchovies, the smallest variety, grow at a slower rate. Anchovies feed on small plankton, chiefly on crustaceans of various families. Fish eggs have also been found in their stomachs. The anchovies are prey to many predatory fishes and marine birds.

The chief countries fishing anchovies include Russia, Spain, Italy, Turkey, the Balkan Countries, Greece, France, and Portugal. Anchovies are caught with drift-nets, baskets, ring-nets, and trawl nets. They are usually marketed in the salted form, in which the head and insides are removed. After an aging period of 4 to 18 months, during which time the flavor improves, the anchovies are ready for market. Part of the catch is worked into filets and conserved in oil. Anchovies are also used in the preparation of pastes and sauces.

Anchovetas. Anchovy fishing is also carried on in other parts of the world, as in the northern Pacific Ocean, on the coast of South Africa, and off Australia. The greatest fishing intensity on a single anchovy species has been carried on since the early 1950s off the coast of Peru and Chile. This species is *Engraulis ringens*, known as *anchoveta* in Spanish. The fish reaches a length of about 5.5 inches (14 centimeters). This is the most important fish in the diet of giant flocks of cormorants, pelicans, gannets, and other marine birds, which breed by the millions on the islands off western South America. The dung of these birds forms the basis of the guano industry. The anchoveta is also eaten by many other animals, including sea lions, dolphins, and predatory fishes.

As of the early 1980s, one of the world's largest fish meal and oil industries was based on these anchovetas. The anchovetas live in the Peruvian Current, ranging from central Chile (37°04′S) all along the Peruvian coast to Cabo Blanco (04°15′S) in a belt close to the shore and extending 30 miles (56 kilometers) out during the summer and 120 miles (222 kilometers) during the winter. The anchovetas spawn in both winter and summer, but with much more intensity and duration during the summer. They reach sexual maturity when approximately 1 year old when they are about 4.7 inches (12 centimeters) long. They can produce about 9000 eggs during several spawnings in the same season. It has been indicated that *E. ringens* spawns from 94°15′S to the south, and that the young anchovetas reach 3.1 and 3.5 inches (8 to 9 centimeters) in length at an age of about 6 months, being then recruited to the fishery. The diet of the species consists of 1% zooplankton and 99% phytoplankton.

The principal fishing gear is the purse-seine net. Most of the catch is reduced to fish meal and oil. Since guano is formed from anchovetas, the proper management of the anchoveta resource is of concern on 2 counts–the fish and the guano. The maximum sustainable catch of anchovetas from fishing and by birds has been estimated at about 10 million tons per year. It has been estimated that the catch by birds approximates 2.5 million tons per year. The Peruvian government has engaged in restricting the anchoveta catch to preserve a satisfactory balance between these two important resources.

Systematic studies of fish larvae have revealed that, with the decline of sardine population, its very close competitor, the anchovy has increased in abundance. It has been estimated by scientists of the California Cooperative Fishery Investigation organization that there exists off California and Baja California a standing stock of some 4 million tons of anchovies, enough to sustain a harvest of perhaps a million tons per year or more. It is believed that some reduction of the anchovy population might, at the same time, accelerate recovery of the heavily depleted sardine population.

See also **Fishes.**

ANCILLARY STATISTIC. In cases where no sufficient statistic exists, it is sometimes possible to find a set of statistics which provide no information on the parameter concerned, but which together with a suitable estimator exhaust all the information in the sample. Such statistics are called ancillary statistics; they provide information, not on the parameter, but on the accuracy with which it is estimated.

ANDALUSITE. An aluminum silicate corresponding to the formula Al_2SiO_5, and is one of a three-member polymorphous group consisting of andalusite, sillimanite, and kyanite. Andalusite occurs in contact-metamorphic shales, and in rocks of regional metamorphic origin in association with sillimanite and kyanite. Andalusite crystallizes in the orthorhombic system, developing coarse prisms of approximately square cross section, but may be massive or granular. It shows a distinct cleavage parallel to the prism; hardness 6.5–7.5; sp. gr., 3.13–3.16; vitreous luster; colorless to white, gray, brown, greenish, or reddish; streak, white; transparent to opaque.

This mineral is named for its original locality, Andalusia, Spain. A variety of andalusite, chiastolite, has carbonaceous impurities so oriented that they produce a cross or a tesselated figure at right angles to the prism. Chiastolite comes from the Greek word meaning a cross. Localities are the Urals, the Alps, the Tyrol, the Pyrenees, Australia, and Brazil; in the United States, at Standish, Maine; Sterling and Lancaster, Massachusetts; Delaware County, Pennsylvania; and Madera County, California.

When clear it is used as a gem, and it has also been used to manufacture porcelain for spark plugs.

AND (Circuit). A computer logical decision element which provides an output if and only if all the input functions are satisfied. A three-variable AND element is shown in Fig. 1. The function F is a binary 1 if, and only if, A and B and C are all 1's. When any of the input functions is 0, the output function is 0. This may be represented in Boolean algebra by $F = A \cdot B \cdot C$, or $F = ABC$. Diode and transistor-circuit schematics for the two-variable AND function are shown in Fig. 2. In modern integrated circuits, the function of the two transistors or diodes may be fabricated as a single active device. In the diode AND circuit, output F is positive only when both inputs A and B are positive. If one or both inputs are negative, one or both diodes will be forward-biased and the output will be negative. The transistor AND circuit operates in a similar manner, i.e., if an input is negative, the associated transistor will be conducting and the output will be negative.

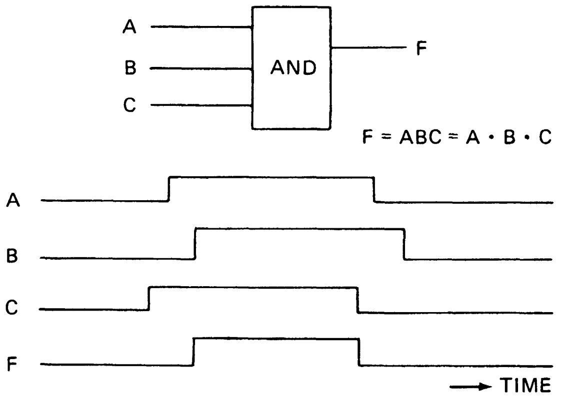
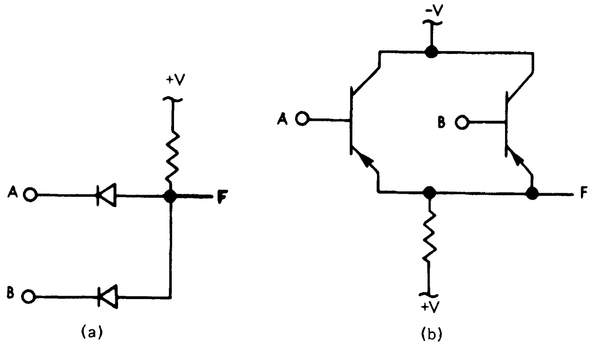


Fig. 1. AND circuit.



(a) (b)

Fig. 2. (a) Diode-type AND circuit; (b) transistor-type AND circuit.

Generally referred to as "fan in," the maximum number of input functions for which a given circuit configuration is capable is determined by the leakage current of the active element. Termed "fan out," the number of circuits which can be driven by the output is a function of current that can be supplied by the AND circuit.

Thomas J. Harrison, International Business Machines Corporation, Boca Raton, Florida.

ANDESITE. A term originally applied to a porphyritic lava from the Andes Mountains by Leopold Van Buch. In modern terminology andesite is an extrusive igneous rock, the surface equivalent of diorite. In other words, it is composed chiefly of plagioclase, corresponding in chemical composition to oligoclase or andesine together, with biotite, hornblende, or pyroxene in varying quantities.

Andesites are of rather widespread occurrence, being found in the Rocky Mountains, California. Alaska, South America, and in many other localities.

ANDRADITE. Calcium-iron garnet.

ANDROGENESIS. The development of an egg after the entry of the male germ cell without the participation of the egg nucleus.

ANDROGENS. The relation between the testis and the male secondary sex characteristics has long been known. The evidence that a chemical substance present in the testis could elicit androgenic effects was not achieved until 1908 by Walker who prepared an aqueous glycerol extract of bull testis tissue that caused growth of the capon's comb. More active extracts from bull testes were prepared in 1927 by McGee and Koch by using organic solvents. The extracts were assayed quantitatively by measuring the increase in area of the capon's comb. The discovery of androgenic activity in urine made possible the isolation of the first biologically active crystalline androgens by Butenandt et al. in 1931–1934. Androsterone and dehydroisoandrosterone were isolated from male urine. In 1935, David et al. isolated a crystalline hormone from bull testis extract. This possessed a higher biological activity than either androsterone or dehydroisoandrosterone. It was named testosterone. Testosterone was also prepared from cholesterol within months of its isolation from testular extract. See Fig. 1.

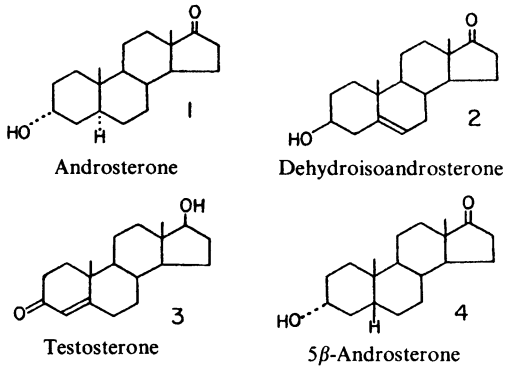


Fig. 1. Androsterone and related hormones.

As pointed out by Liddle and Melmon (1974), adrenal androgen production also carried out in the *zona fasciculata* and in the *zona reticularis*, varies greatly at different stages of life. The fetus makes significant amounts of adrenal androgen, whereas the child makes very little. Beginning with puberty, adrenal androgen production increases, reaches a peak in early adulthood, and then declines to rather low levels beyond age 50. On the other hand, the secretion of ACTH, the only known control of adrenal androgen biosynthesis, shows no age-related fluctuations. The full regulation of adrenal androgen production is not understood. Adrenal androgens are relatively weak, but some serve as precursors for hepatic conversion to testosterone. Hyperfunction of this pathway in the female may lead to significant masculinization. See also **Adrenal Glands.**

The androgens stimulate the development of the male secondary structures, such as the penis, scrotum, seminal vesicles, prostate gland, vas deferens and epididymis. The deepening of the voice, the growth of pubic, axillary, body, and facial hair, as well as the development of the characteristic musculature of the human male, are also under the influence of testosterone. If the testes fail to develop or are removed prior to puberty, these changes do not occur. Thus, testosterone is essential for reproductive function of the male.

The adrenal cortex produces hydroisoandrosterone which is found in blood and urine largely conjugated as the sulfate ester. The amounts of androgen secreted by the normal adrenal cortex are insufficient to maintain reproductive function in the male. The normal human ovary and placenta also produce small amounts of androgenic steroids that serve as precursors for the estrogens in these tissues. In the human, little testosterone is excreted into the urine and virtually none into the feces. The principal metabolic transformation products are androsterone and 5β-androsterone, with small amounts of other reduced compounds.

These substances are excreted in the urine in the form of esters with sulfuric acid or glycosides with glucuronic acid.

Like all other classes of steroid hormones, the androgens are synthesized from acetyl coenzyme A *via* mevalonic acid, isopentenyl pyrophosphate, farnesyl pyrophosphate, squalene, lanosterol, and cholesterol. Enzyme systems in the testis then catalyze the cleavage of the sidechain of cholesterol to pregnenolone which can give rise to testosterone by the two pathways shown in Fig. 2.

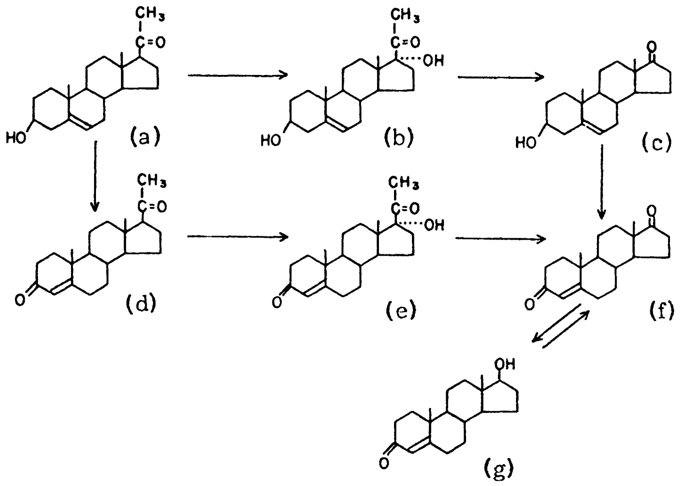


Fig. 2. Biosynthesis of testosterone: (a) Pregnenolone; (b) 17-hydroxy pregnenoline; (c) dehydroisoandrosterone; (d) progesterone; (e) 17-hydroxyprogesterone; (f) androstenedione; (g) testosterone.

Testosterone is formed by the interstitial or Leydig cells of the testes which develop under the influence of gonadotrophic hormones discharged into the bloodstream by the anterior pituitary gland. In pituitary insufficiency, this hormonal stimulus is lacking and, as a consequence, the Leydig cells do not secrete testosterone. In such instances, the male secondary sex characteristics fail to develop. However, interstitial cell tumors may occur, leading to excessive androgen production and precocious puberty. In women, tumors or excessive function of the adrenal cortex and, rarely, of the ovary, result in the production of large amounts of androgens with associated virilization.

Acne vulgaris, a chronic skin disorder, is related to androgen production. The postpubescent development of the sebaceous glands and the onset of acne are dependent upon the presence of androgens, but are not related to testosterone blood levels. Sansone et al. (1971) observed that the hypothesis of increased end-organ sensitivity, which is supported by the heightened ability of skin with acne to metabolize testosterone, may explain the lack of correlation between levels of circulating androgens and occurrence and severity of the disease.

There is androgen involvement in polycystic ovary syndrome (PCO) and hyperthecosis. In 1935, Stein and Leventhal defined a condition with hirsutism, secondary amenorrhea, and enlarged ovaries—a syndrome now referred to as PCO. In this condition, the female usually shows signs of androgen excess, including increased body hair, but true virilism, with balding and deepening of the voice, is less common. Usually, one or both ovaries are enlarged. In many patients the ovaries are cystic, with thickened capsules, yet not palpably enlarged. It has been postulated that the development of PCO commences when luteinizing hormone (LH) triggers an increase in ovarian androgen, which is converted to estrogen, causing estrogen levels (particularly estrone) to increase. This is followed by an anterior pituitary response to luteinizing hormone-releasing hormone (LRH). This completes the cycle by creating exaggerated pulsatile but surgeless LH levels. The initiating lesion remains obscure. Wedge resection of one or both ovaries has largely been replaced by the administration of the antiestrogen clomiphene. This was the first drug known to trigger ovulation in women. Patients with anovulation arising from PCO are treated with the drug primarily when fertility is desired. The hirsutism found with PCO has been difficult to treat. Wood and Boronow (1976) reported that long intervals of anovulation, such as occur in PCO, may be a prelude to endometrial carcinoma. As postulated, the link may either be continuous exposure of the endometrium to estrogen unopposed by progesterone. There is also the postulation that estrone, the estrogen that appears to be high in this disorder, may be a causative factor. A carcinogenic role has been alleged for estrone.

In a somewhat related disorder, hyperthecosis, androgen excess tends to be greater. In this condition, there is prominent luteinization of the theca, whereas the cystic development and capsular thickening of the PCO are absent. Ovarian tumors making androgen can produce the features of PCO, but they tend to have a course with sharper onset and clearer progression.

Androgens also have been used in the management and treatment of agnogenic myeloid metaplasia, aplastic anemia, breast cancer, hereditary angiodema, osteoporosis, paroxysmal nocturnal hemoglobinuria, and sideroblastic anemia.

Oral androgens as may be used in hormone therapy for the management of metastatic breast cancer are effective in about 30% of women regardless of age, but virilizing doses of the hormone are usually required. The oral androgens carry the additional risk of toxic hepatitis.

Androgens may explain some of the differences between heart diseases of males and females. See **Heart and Circulatory System (Human).**

Additional Reading

Austin, C. R., and R. V. Short, Eds.: "Hormonal Control of Reproduction," 2nd Ed., Cambridge University Press, New York, 1984.

Barbieri, R. L. and K. Schiff, Eds.: "Reproductive Endocrine Therapeutics;" Alan R. Liss, New York, 1988.

Besser, G. M., and W. J. Jeffcoate: "Endocrine and Metabolic Disease: Adrenal Diseases," *Br. Med. J.*, **1**, 448 (1976).

Eldar-Geva, T. et al.: "Secondary Biosynthetic Defects in Women with Late-Onset Congenital Adrenal Hyperplasia," *New Eng. J. Med.*, 855 (September 27, 1990).

Greydamus, D. E. and R. B. Shearin: "Adolescent Sexuality and Gynecology," Lea and Febiger, Philadelphia, Pennsylvania, 1990.

Griffin, J. E.: "Androgen Resistance—The Clinical and Molecular Spectrum," *New Eng. J. Med.*, 611 (February 27, 1992).

Kidd, K. K.: "The Search for the Ultimate Cause of Maleness," *N. Eng. J. Med.*, 260–261 (July 23, 1985).

Naftolin, F.: "Understanding the Bases of Sex Differences," *Science*, **211**, 1263–1264 (1981).

Smith, E. L., et al., Eds.: "Principles of Biochemistry: Mammalian Biochemistry," 7th Ed., McGraw-Hill, New York, 1983.

Williams, R. H., Ed.: "Textbook of Endocrinology," 6th Ed., W. B. Saunders, Philadelphia, 1981.

Wilson, J. D., et al.: "The Hormonal Control of Sexual Development," *Science*, **211**, 1278–1284 (1981).

Yen, S. S. C.: "The Polycistic Ovary Syndrome," *Clin. Endocrinol. (Oxford)*, **12**, 177 (1980).

Yen, S. S. C. and R. B. Jaffe: "Reproductive Endocrinology: Physiology, Pathophysiology and Clinical Management," W. B. Saunders, Philadelphia, Pennsylvania, 1991.

ANDROMEDA. The brighter stars of this constellation make an almost straight line between the constellations of Perseus and Pegasus. The most famous feature of the constellation is the great spiral galaxy. This is the only spiral actually visible to the naked eye, and may be distinguished as a faint blur against a moonless sky close to the faintest star in the constellation. (See map accompanying entry on **Constellations.**) The distance of this spiral from the earth is about 1.84×10^6 light years.

The bright star in Andromeda closest on the map to Perseus was called Almach by the Arabs. It is a double star, and one of the most beautiful in the sky when viewed with a small telescope. One component is a brilliant orange and the other a striking emerald color. Careful examination with a large telescope shows the green component to be also a double star.

ANECHOIC ROOM. This term means literally a room without echoes, which is actually a room in which sound reflections from the boundary surfaces have been reduced to a negligible amount.

ANEMIAS. The anemias comprise one of the major groups of diseases involving the blood and, in particular, the erythrocytes (red cells) and their hemoglobin, that is, the impairment of the blood's oxygen delivery system. Anemias may arise from (1) blood loss; (2) disorders of iron metabolism; (3) defects in erythrocyte production; and (4) hemolysis— destruction of erythrocytes.

A male weighing about 150 pounds (70 kilograms) will have just over 4 grams of iron in his body, with 61.7% in the form of hemoglobin; 3.5% (myoglobin); 0.2% (heme enzymes); and 34.6% in iron stores (transferrin, hemosiderin, ferritin). In a woman, there is a marked difference in the distribution of the iron. A female weighing about 132 pounds (60 kilograms) will have about 2.3 grams of iron in her body, with 81.4% in the form of hemoglobin; 5.4% (myoglobin); 0.3% (heme enzymes); and 12.9% in iron stores. The small quantity in iron stores is represented by blood loss in menstruation and in pregnancy and lactation.

Blood Loss and Iron Deficiency Anemia

Blood loss is classified as acute or chronic. In instances of blood loss from injuries, this may be immediately obvious to both patient and physician. Where there is massive acute blood loss, shock and death will occur if replacement therapy is not commenced in very short order. Where blood losses approximate one liter, there will be symptoms of incipient or overt shock. Symptoms will progress into shock, depending upon the severity of further losses. Also, in accidents, there may be deep tissue bleeding where blood losses may not be immediately apparent or measurable. For example, in the case of a fractured pelvis, where there is hemorrhage into the thigh and pelvic region, a liter or more of blood may be lost and not immediately detectable. In all severe cases of bleeding, infusions with colloid and electrolyte solutions will be commenced as soon as possible and before blood typing and cross-matching procedures can be completed. See **Blood.** Frequently, when available, a preparation known as *plasmanate* will be used. This preparation contains human plasma and a small quantity of albumin. Contrasted with dextran infusions, plasmanate does not cause platelet functional defects or red cells to aggregate. Concurrently, equal volumes of a saline-based electrolyte solution will be given. Once the steps required to restore plasma volume and electrolytes have been taken, the physician will tackle the problem of replenishing the red blood cells.

Chronic blood loss is frequently attributed to gastrointestinal tract bleeding and, in women of child-bearing age, to blood losses from the genital tract. An average woman will lose from 30 to 60 milliliters of blood per month through menstruation. During pregnancy, there is division of iron from the mother to the fetus. There are further losses during parturition and lactation. A mother may lose from 700 to 900 milligrams of iron in this way. All of these factors may be contributory to *iron deficiency anemia*. Other factors may include disturbances in the absorption of iron and deficiency of dietary iron. Less frequently, iron deficiency anemia will be a result of intravascular hemolysis resulting in iron loss through the urine (*hemoglobinuria* and *hemosiderinuria*); or even a result of loss of blood to the lungs in an uncommon condition known as idiopathic pulmonary hemosiderosis.

Iron replacement therapy involves the oral or intravenous (depending upon severity and patient reactions) administration of iron-containing compounds (commonly ferrous sulfate). The therapy is usually commenced at low levels and gradually increased so that gastrointestinal symptoms may be avoided.

In determining iron replacement requirements, numerous factors must be considered. There is no single laboratory value that defines anemia. For example, a person may lose nearly a liter of blood and show signs of impending shock from a bleeding peptic ulcer, yet samples of peripheral blood may show normal hemoglobin. This occurs because dilution of the blood (to restore loss of volume) does not occur for about 72 hours. As guidelines, the following stages of iron deficiency are used by some specialists in the field: *Normal*, a hemoglobin level of 13–15 grams per 100 milliliters; *iron deficiency without anemia*, same values; *iron deficiency with mild anemia*, a hemoglobin level of 9-10 grams per 100 milliliters; *severe iron deficiency with severe anemia*, a hemoglobin level of 6-7 grams per 100 milliliters.

Anemias from Red Blood Cell Production Disorders (Erythropoiesis)

Although anemia may result from defects in red blood cell production alone, frequently one of these disorders will be accompanied by other factors (hemolysis or blood loss) which exacerbate the anemia. Sometimes, there are also associated depression of platelet and white blood cell counts.

As mentioned in entry on **Blood,** the erythrocytes are produced in bone marrow. In what is sometimes called the *anemia of chronic disorders*, the bone marrow appears to be normal. Examination will show that there is a normal ratio of myeloid cells (precursors of the erythroid cells) in the marrow and of the erythroid cells produced. This is classified as a mild anemia and is usually presented by patients who have chronic inflammatory, infectious, or neoplastic (presence of tumors) disease. Diagnosis of the underlying condition is often difficult. Anemia is frequently found in severe renal (kidney) disease. As described in the entry on **Blood,** a chemical messenger (*erythropoietin*) is released by the kidney to signal the rate of erythrocyte generation required. Dysfunction of this system can be a causative factor. Where renal disease is accompanied by chronic uremia (blood in urine) resulting from gastrointestinal bleeding, iron deficiency will add to the complications.

Also not directly marrow related are anemias resulting from starvation, such as anorexia nervosa or protein deficiency. These conditions may arise even though normal folate and vitamin B12 levels are maintained. Therapy is improvement of the diet. See **Anorexia.** Reduced red blood cell production not directly involving the marrow may also be caused by certain drugs, such as alcohol (which interferes with metabolism of folate and iron), chloramphenicol, and arsenic, among others, However, the latter drugs also can affect the marrow.

Aplastic Anemia. In this anemia, there is partial or nearly complete failure of the marrow to produce new red blood cells. This condition may arise from several causes, including ionizing irradiation, a number of chemotherapeutic drugs, as well as several diseases. There are also instances of idiopathic aplastic anemia which may be due to defective behavior of the stem cells. Benzene also has been implicated as a causative factor. Vigorous inhalation of some organic vapors (glue sniffing) can induce fatal aplastic anemia.

Among drugs that are frequently implicated in aplastic anemia are various *alkylating agents*, such as melphalan, cyclophosphamide, chlorambucil, and bisulfan; *antimetabolites*, including azathioprine, 6-mercaptopurine, 6-thioguanine, and methotrexate; and various *antitumor agents*, such as vinca alkaloids (vinblastine, vincristine), anthracyclines (daunorubicin, doxorubicin), among others. Drugs that occasionally cause aplastic anemia include arsenic, chloramphenicol, gold compounds, mesantoin, phenylbutazone, quinacrine, sulfonamides, and trimethadione. Diseases associated with aplastic anemia include viral hepatitis and paroxysmal nocturnal hemoglobinuria.

Biopsy may be required to determine aplasia of the marrow. In some cases, it may be found that the marrow has been replaced by tumors or fibrosis. Aplastic anemia has been treated with corticosteroids and splenectomy has been done, but their effectiveness has not been well documented. Aplastic anemia is a very serious disease and not always effectively treated, particularly if there are few if any surviving pluripotent stem cells. A marrow transplant may be considered. These are not always successful because of graft rejection, but the procedure may be the only remaining way of saving some patients' lives.

Megaloblastic Anemias. Commonly called *pernicious anemia*, this disorder is caused by vitamin B_{12} and folic acid deficiencies. In megaloblastic anemia, several features of the interactions between vitamin B_{12} and folic acid coenzymes are critical. See also **Folic Acid;** and **Vitamin B_{12}.** Because neither of these substances are produced by humans in adequate amounts, they must be absorbed from a good diet. Factors which cause vitamin B_{12} deficiency include: (1) Inadequate diet, particularly resulting from strict vegetarianism; (2) inadequate absorption, such as from gastric abnormalities with deficient or defective intrinsic factor, small bowel disease, and pancreatic insufficiency; (3) interference with vitamin B_{12} absorption as caused by fish tapeworm and certain drugs, such as neomycin, colchicine, para-aminosalicylic acid, and ethanol; and (4) rare congenital disorders, such as transcobalamin-II deficiency or defective intrinsic factor production. Factors which cause folic acid deficiency include: (1) Inadequate intake, as in nutritional deficiencies and alcoholism; (2) relatively inadequate intake, as may occur during pregnancy, severe hemolysis, and chronic hemodialysis; (3) inadequate absorption, as occurs in tropical sprue, Crohn's disease, lymphoma or amyloidosis of small bowel, diabetic enteropathy, and intestinal resections or diversions; and (4) interference with folic acid metabolism, as may be precipitated by drugs blocking the action of dihydrofolate reductase (methotrexate, trimethoprim, and pyrimethamine), and by other drugs, the exact mechanisms of which are not known—phenytoin, ethanol, antituberculosis drugs, and possibly oral contraceptives.

Pernicious anemia usually does not occur before middle life. It results from the disappearance of the *intrinsic factor* and with it, hydrochloric acid from gastric juices. Upon progression of the disease, certain changes can occur in the spinal cord which result in weakness and numbness of the limbs and ultimately a full loss of ability to control them. Added to weakness and pallor, the symptoms may include loss of appetite, diarrhea, nausea, sore tongue, and yellow pigmentation of the skin. Until 1926, no treatment was known. In that year Minot and Murphy, American physicians, introduced the use of dietary liver as a specific treatment for patients suffering with pernicious anemia. For this work, they received the Nobel prize in 1934.

Upon diagnosis of the disease, vitamin replacement therapy should be commenced immediately. Where the patient is symptomatic from severe anemia, packed red cells can be transferred very slowly to avoid precipitating or aggravating congestive heart failure. This will usually produce a 25% increase in oxygen-carrying capacity of the blood within a short period. Large, weekly doses of parenteral vitamin B_{12} are administered for several weeks, after which these may be scheduled on a monthly basis. Monthly doses may be required for the remainder of life. The physician will also encourage good dietary practice. Depending upon diagnosis, oral administration of folic acid may be indicated.

Sideroblastic Anemias. These comprise a heterogeneous group of disorders characterized by anemia and ineffective erythropoiesis.

Hemolytic Anemias

The anemias which result from increased red blood cell destruction are termed *hemolytic anemias*; they may be *normocytic* (red cells are of normal size) or *macrocytic* (red cells are larger than usual). Hemolysis may be caused by several differing conditions. In anemias caused by hemolysis, the breakdown of red cells releases large amounts of hemoglobin end products into the plasma. These substances are converted by the liver into a number of other pigments, most of which are excreted in the bile. If production of bile pigments is excessive, some appear in body tissues, giving rise to a yellow appearance of the skin and the whites of the eyes. This condition is termed *jaundice* and is a symptom of the hemolytic anemias. The red cells are abnormally fragile and rupture easily in *hemolytic jaundice*. Thus, they are broken down by the spleen more rapidly than is usual. Such cells, without interference of the spleen, are able to function normally even though they are fragile. Thus, in some cases of hemolytic jaundice, the spleen is removed as a means of preventing too-rapid destruction of these cells. Some kinds of hemolytic anemia are inherited. Other types are acquired and may be associated with various systemic diseases. A number of drugs, physical and chemical agents, and vegetable and animal poisons have been suspect as causes. Corticosteroid therapy can be beneficial in some of these cases. Where there is no response to treatment, removal of the spleen is indicated.

Sickle Cell Anemia. This disease is caused by hemoglobin S that is inherited as a Mendelian dominant characteristic. It occurs as the *sickle trait* in 8–10% of black persons in the United States. In persons with sickle trait, the hemoglobin S concentration is less than 50% and, with rare exceptions, there are no symptoms. However, in *sickle cell anemia*, 70–98% of the hemoglobin is of the S type, leading to severe disease. The distribution of hemoglobin S in localities with a high incidence of malaria has suggested to some investigators that the sickle trait may provide some advantage to those who possess it. Studies have shown that persons with the sickle trait (but *not* with sickle cell anemia) are relatively resistant to the serious effects of falciparum malaria. It has been presumed that, in patients with the sickle trait, the parasitized cell "sickles" and thus is removed from the circulation in a sequence of events that breaks the parasite's life cycle. See **Malaria.**

Defective hemoglobin results in misshapen red cells and an inability of the blood to carry sufficient oxygen, thus producing anemia. Symptoms, as in the instances of other anemias, include general weakness and, in severe cases, headache, nausea, vomiting, fever, jaundice, and muscular and joint pains. The *sickle crisis*, which occurs in this disease, is a painful and dramatic expression of vascular occlusion. The initiating factor in the sickle crisis is not fully understood. Episodes of fever are known to predispose a patient to crisis. Studies in Ghana have shown that at the beginning of the malaria season, there are sharp increases in the incidence of sickle crisis. The traditional handling of the painful sickle crisis includes rest, the administration of drugs to relieve pain, and, if the patient is demonstrably acidotic, the administration of sodium bicarbonate in a 5% dextrose-water solution, normal saline solution, or half-normal saline solution. The bicarbonate solution is infused for a period of about 20 hours. During a crisis, the prevention of infection and other complications is very important. Past trials of oxygen therapy and urea therapy have not proved beneficial. Where a sickle crisis is of extreme severity, exchange transfusions may be indicated. As blood components are administered, part of the patient's blood will be drawn off. This procedure may be repeated three or more times, at the fastest rate permissible. Particularly in pregnant women, exchange transfusions are commenced at the beginning of the third trimester to avoid complications in pregnancy, which may include fetal death. Transfusions may be carried out at weekly intervals. Such a program also is sometimes used prior to surgery.

Persons with sickle cell anemia should be warned about the additional dangers of high altitude and dehydration. Genetic counseling to prospective parents is universally recommended among authorities.

During the 1980s, much research was directed toward a better understanding of sickle cell anemia and the development of improved methods of treatment and possible ultimate prevention.

In the references cited, Marx (1984) describes how parvovirus B19 causes a shutdown of red blood cell production in children with sickle cell anemia. Although this virus was discovered in human blood in 1975, it previously had not been associated directly with any specific disease. Marx (1985) reports of a new sickle cell test. This prenatal diagnostic test involves novel techniques for amplifying and analyzing specific DNA segments. Sickle cell anemia is a hereditary disease caused by the alteration of a single nucleotide in the beta-chain gene, which encodes one of the two proteins of the adult hemoglobin molecule. Individuals who inherit two copies of the mutant gene get sickle cell disease. Persons who inherit just one copy do not have the full-blown disease, but can pass the defective gene on to their children. Chien (1984) reports on microcirculation studies of sickle cell patients that provide insights to the rheological characteristics of sickle cells. In homozygous sickle-cell disease, deoxygenation of red cells causes the polymerization of hemoglobin S and transformation of the intracellular fluid into a viscoelastic gel. This process is most prominent in particularly dense red cells that have a high concentration of hemoglobin S. The membrane of these dense red cells also has elevated viscoelastic moduli, probably as a result of mechanical stresses exerted during sickle-unsickle cycles. Therefore, the densest cells are most rigid and have the greatest effect on resistance in blood flow through the microcirculation. Studies of the retina of patients with sickle-cell disease and the microcirculation of animals receiving human sickle-cell red cells have demonstrated that such red cells tend to be trapped at the entrance of narrow capillaries. Although only a small fraction of sickle-cell red cells have an abnormal rigidity at arterial oxygen saturation, their effects on microcirculatory flow may be magnified by their slow transit. Because the slowly moving rigid cells remain at the precapillary resistance sites for longer periods than cells with normal deformability, they may disproportionately occupy most of these sites and exert a much greater effect on flow resistance than would be expected on the basis of their low percentage in the red-cell population.

Nagel et al. (1985) report on an interesting study of the types of sickle-cell anemia found in the black population of Africa, in which three different haplotypes were found in three different and specific geographical areas—the Senegal-type, the Benin-type, and the Bantutype. It has been established for many years that black patients with sickle cell anemia in North America vary in the hematologic and clinical features of their disease. The variations generally can be attributed to the different geographical areas of origin on the African continent.

Additional Reading

Bick, D., et al.: "Brief Report: Intragenic Deletion of the KALIG-1 Gene in Kallmann's Syndrome," *N. Eng. J. Med.*, 1752 (June 25, 1992).

Browne, P. V., et al: "Donor-Cell Leukemia after Bone Marrow Transplantation for Severe Aplastic Anemia," *N. Eng. J. Med.*, 710 (September 5, 1991).

Caviness, V. S., Jr.: "Kallmann's Syndrome—Beyond 'Migration'," *N. Eng. J. Med.*, 1775 (June 25, 1992).

Charache, S.: "Problems in Transfusion Therapy," *N. Eng. J. Med.*, **1666** (June 7, 1990).

Frickhoffen, N., et al.: "Treatment of Aplastic Anemia with Antilymphocyte Globulin and Methylprednisolone with or without Cyclosporine," *N. Eng. J. Med.*, 1297 (May 9, 1991).

Goldberg, M. A., et al.: "Treatment of Sickle Cell Anemia with Hydroxyurea and Erythropoietin," *N. Eng. J. Med.*, 366 (August 6, 1990).

Golde, D. W.: "The Stem Cell," *Sci. Amer.*, 86 (December 1991).

Kodish, E., et al.: "Bone Marrow Transplantation for Sickle Cell Disease—A Study of Parents' Decisions," *N. Eng. J. Med.*, 1349 (November 7, 1991).

Ludwig, H., et al.: "Erythropoietin Treatment of Anemia Associated with Multiple Myeloma," *N. Eng. J. Med.*, 1693 (June 14, 1990).

Miller, C. B. et al: "Decreased Erythropoietin Response in Patients with the Anemia of Cancer," *N. Eng. J. Med.*, 1689 (June 14, 1990).

Milner, P. F., et al.: "Sickle Cell Disease as a Cause of Osteonecrosis of the Femoral Head," *N. Eng. J. Med.*, 1476 (November 21, 1991).

Moore, M. A. S. and H. Castro-Malaspina: "Immunosuppression in Aplastic Anemia—Postponing the Inevitable," *N. Eng. J. Med.*, 1358 (May 9, 1991).

Platt, O. S., et al.: "Pain in Sickle Cell Disease—Rates and Risk Factors," *N. Eng. J. Med.*, 11 (July 4, 1991).

Vichinsky, E., et al.: "Alloimmunization in Sickle Cell Anemia and Transfusion of Racially Unmatched Blood," *N. Eng. J. Med.*, 1617 (June 7, 1990).

Ware, R. E., Hall, S. E. and W. F. Rosse: "Paroxysmal Nocturnal Hemoglobinuria with Onset in Childhood and Adolescence," *N. Eng. J. Med.*, 991 (October 3, 1991).

ANEMOMETER. See **Wind and Air Velocity Measurements.**

ANEMOTAXIS. Orientation of an insect to an air current as an inflight mechanism in seeking out a distant odor source. Some scientists now claim that insects also can follow an airborne odor trail in still air.

ANESTRUS. A period in which there is lack of heat (estrus) in the female animal, thus precluding breeding during that period. The periods of sexual noninterest in the female that normally occur between regular heat periods generally are not referred to as anestrus. Rather, *anestrus* applies to those factors that break up the normal cycles of heat. For example, there is lactational anestrus, which is a period following the birth of the young; or there is seasonal anestrus, which is a perfectly natural condition with some animals. For example, in some regions, seasonal anestrus occurs in sheep during late-spring and early-to-midsummer months. Abnormal anestrus is always of concern to cattle, sheep, and swine producers because of the reduction in production of calves, piglets, and lambs during any given period.

ANEURYSM. A sac or pouch filled with blood which protrudes from the wall of an artery, a vein, or the heart. In a *true* aneurysm, the wall of the sac consists of at least one of the layers of tissue that make up the wall of the blood vessel. *False* aneurysms exist when all of the layers of the artery have ruptured, but the blood is still retained by the surrounding tissues. Occasionally, an artery and vein may be connected in such a manner that a continuing flow of blood passes from the artery to the vein. Such arteriovenous communications result from wounds, aneurysms, or congenital connections between the vessels.

Aneurysms may occur in any artery. The most common site is the large artery leading from the heart (the *aorta*). Small aneurysms sometimes develop in blood vessels as the result of injuries. The rupture of even a small aneurysm in the brain, heart, or other vital organ can be fatal.

Presently, the most common cause of aortic aneurysm is atherosclerosis known as hardening of the arteries. Also, an injury to an arterial wall can leave it so weakened that an aneurysm eventually may occur. Infected (*mycotic*) aneurysms result from destruction of arterial walls by infectious agents. Typical of diseases that may leave a weakened blood vessel are pneumonia, streptococcal infections, and gonorrhea. Now very rare, aneurysms of the aorta formerly were common as the result of previous syphilitic infection.

Although an aneurysm may exist at any site along the aorta, the most common aneurysms are in the abdominal aorta. Symptoms depend upon size of the sac and parts of the body upon which it exerts pressure. The sacs frequently become large, sometimes larger than an orange and can cause severe crowding of the chest or abdominal cavity. Bulging of the area of the collar bone can result when the aneurysm is near the top of the aorta. Sometimes aneurysms are painful, the pain usually located in the center of the chest, or may radiate into the arms. Difficulty in breathing (*dyspnea*) is another common symptom. This results from pressure on the windpipe or smaller air passages leading to the lungs. Aneurysms may cause headaches, abdominal distress, or swelling in various parts of the body. Many aneurysms are symptomless, however, and may not be discovered except as the result of an x-ray examination for some other purpose. Angiography, in which contrast material is introduced into the blood vessels, followed by x-ray examination, is a valuable diagnostic aid.

Although many aneurysms remain small and never require treatment, generally they tend to become larger and may progress to rupture. The major treatment is complete surgical excision of the aneurysm with subsequent restoration of circulation with an implanted, pleated artificial graft. Where the artery is small, the graft may be a segment of vein from the patient. Risk of rupture and development of more hazardous problems rise with a delay in surgical procedure.

For smaller blood vessels which supply nonvital areas, aneurysms may be tied off so that blood no longer passes through the distended area. Other arteries assume the work of the vessel that is closed-off.

Most cerebral aneurysms occur in or near the circle of Willis: 30% of cases are found in the anterior communicating artery; 25% in the carotid artery; 25% in the posterior communicating artery; 2% in the basilar artery; and 2% in the vertebral artery. Several other sites are less frequently involved. See **Nervous System and The Brain.** Sometimes large aneurysms will produce focal neurologic signs that lead to a correct diagnosis prior to rupture. These aneurysms may occur at any age, but middle-aged persons are most prone to them. See also **Cerebrovascular Diseases.**

ANGEL. A radar echo caused by a physical phenomenon not discernible to the eye. Angels are usually coherent echoes, whose phase and amplitude at a given range remain relatively constant, and are sometimes of great signal strength (up to 40 decibels above the noise level). They have been ascribed to insects flying through the radar beam, but have also been observed under atmospheric conditions that indicate there must be other causes. Studies indicate that a fair proportion of angels are caused by strong temperature and/or moisture gradients such as might be found near the boundaries of bubbles of especially warm or moist air. They frequently occur in shallow layers at or near temperature inversions within the lowest portion of the atmosphere.

ANGELFISHES (*Osteichthyes*). Of the family *Pomacanthinae*, angelfishes have a powerful spine on the lower rear edge of the front gill cover; there is generally no axillary scale at the base of the pectoral fin. The spine distinguishes angelfishes from the closely related butterfly fishes (*Chaetodontinae*) which have no similar spine. Angelfishes are among the most beautiful of fishes and are usually found on tropical reefs. They tend to travel alone or in small groups. Although most angelfishes are much smaller, some reach a length of 2 feet (0.6 meter). These fishes are not only well known to the fishers in their native habitats, but have also become very popular in Europe and North America as aquarium fishes. Their image often is found on aquarium hobbyist organization emblems, since hobbyists have long been attracted to them. See accompanying figure.

The length of the angelfish *Pterophyllum scalare*, when fully developed in its natural habitat, may attain about 10 inches (25 centimeters). This fish inhabits the entire Amazon and its tributaries, resulting in the

Amazon angelfish (*Pterophyllum scalare*). (*Sketch by Glenn D. Considine.*)

development of various geographic races differing in shape as well as coloration. Spawning usually occurs on a strong plant stem, or a large leaf. Both partners clean off the spawning surface beforehand. Both parents also care for the young together. Scalares do not build spawning pits to which the young are brought after hatching. Instead, the young fish are carried in the parents' mouth and suspended either directly on the spawning surface, or some other cleaned area. This is done with the help of fibers projecting from adhesive glands on the top of the heads of the young. Both parents guard the developing eggs and the hatched young.

Species of angelfish include:

Blue Angelfish (*Pomacanthus semicirculatus*), tropical Indo-Pacific.
Blue-faced Angelfish (*Pomacanthus xanthometopon*), tropical Indo-Australian region.
Imperial Angelfish (*Pomacanthus imperator*), tropical Indo-Pacific.
French Angelfish (*Pomacanthus arcuatus*), tropical Atlantic, from the Florida Keys to Brazil.
Potter's Angelfish (*Centropyge potteri*), Hawaiian Islands.
Queen Angelfish (*Holacanthus ciliaris*), southwestern and northeastern Gulf of Mexico to Brazil.
Regal Angelfish (*Pygoplites diacanthus*), central Indo-Pacific (tropical).
Rock Beauty (*Holacanthus tricolor*), Florida to Brazil.

ANGEL SHARK (*Chondrichthyes*). The species *Squatina squatina* is the largest angel shark, with a maximum length of about 8 feet (2.5 meters) and weight of 160 pounds (73 kilograms). Angel sharks are intermediate between sharks and rays. Other species of the family *Squatinidae* are seldom longer than 4 to 5 feet (1.2 to 1.5 meters) at maturity. They prefer inshore waters, rarely frequenting deep waters. One exception is the Atlantic American *Squatina dumeril*, one specimen of which was found at a depth of some 4200 feet (1260 meters). Although angel sharks are not considered dangerous to swimmers, they are of a nasty disposition when out of the water and thus, if hooked, pose a danger to fishermen. They survive in captive waters for only a few weeks or months. The Pacific American *Squatina californica* inhabits waters from Alaska to lower California and like most angel sharks prefers temperate waters.

ANGIOGRAPHY. In making an angiogram, a catheter is inserted into the individual's heart, a radiopaque medium is injected, and x-ray images and motion pictures may be made. These pictures indicate the locations where arteries are blocked and the degree to which the blockage has developed. Motion picture studies show details of heart function. Contrast angiograms are usually indicated when surgery or other therapy is being considered for persons with angina pectoris and those who are recovering from heart attack. In another version of this technique, the heart can be labeled with radioactive tracers which emit gamma rays. Images are made with a scintillation camera that is sensitive to gamma radiation. Through the use of computer techniques, three-dimensional x-ray or gamma-ray pictures of the heart can be created. Scintillation camera images have come into rather wide application as a means of diagnosing heart attacks, particularly where electrocardiograms may not be fully definitive, and where certain other clinical findings, such as serum enzymes, may not confirm heart damage. Technetium-99m, a radionuclide tracer, has a particular affinity for recently damaged heart muscle. After intravenous injection, these tracers find their way to damaged areas and emit gamma rays which are picked up as bright areas in the scintillation camera image. In persons with unstable angina pectoris, where other clinical findings are indefinite, the radioactive tracers tend to collect in the heart. Another approach is to inject radioactive tracer consisting of a monovalent cation, such as potassium-43 or thallium-201. Upon intravenous injection of these materials, the substances lodge in the heart in proportion to blood flow. Portions of the heart that are deprived of blood flow thus appear as blank regions in the scintillation camera image.

In modern medicine, where useful information can be obtained by way of noninvasive techniques, the latter are preferred. Many advancements in non-invasive procedures have been made during the past decade and, to some extent, these have impacted on invasive procedures, such as traditional angiography as just described. Noninvasive techniques are of particular preference in the early stages of diagnosis in the absence of an emergency situation. Such techniques include echocardiography and computed tomography (CT) and nuclear magnetic resonance (NMR).

See **Heart and Circulatory System (Human); Ischemic Heart Disease;** and **Radioisotopes.**

ANGIOMA. A tumor which is composed mainly of blood vessels (hemangioma) or of lymph vessels (lymphangioma). Both forms are ordinarily harmless: in the skin a hemangioma may appear as a disfiguring *naevus* or "port-wine stain"; lymphangiomas generally form soft bulky swellings especially about the neck.

ANGIOSPERMS. The angiosperms represent the most advanced division of the Pteropsida. They are more familiarly known as flowering plants, and the characteristic feature is the flower. The seeds are borne completely enclosed in the ovary tissue of the parent plant.

As a rule the angiosperms are land plants. A few of them have returned to the water as a habitat, but these are obviously reversions to an aqueous life. Tremendous diversity in size is found in this group; some of the so-called duckweeds are spherical masses of cells less than a millimeter in diameter; at the other end of the scale are the giant *Eucalyptus* trees, many of which are over 300 feet (90 meters) high. The variety of form shown in the angiosperms is nearly endless; each of the nearly 200,000 species has a distinct appearance. Some are tiny, herbaceous plants which live but a few weeks; others are giant trees living hundreds of years.

Included in the angiosperms are two types of plants. One, held to be the more primitive type, has a woody stem which has a much more complex structure than that found in the stems of gymnosperms. The other has an herbaceous stem, a form of stem which dies to the ground at the end of the growing season. The internal structure of angiosperm stems is much more specialized than that of gymnosperms. The xylem contains not only tracheids but also vessels and fibers. The vessels are open tubes of considerable length through which water is carried rapidly. The fibers give strength to the stem. In the phloem there are sieve tubes and companion cells, and also numerous fibers. In the woody angiosperms and in many of the herbaceous forms there is a well-developed cambium.

Reproduction in the angiosperms is described under **Flower.** Many of the angiosperms have highly developed mechanisms to ensure pollination and fertilization, and elaborate systems for the protection, dissemination, and germination of the seeds.

The angiosperms are separated into two large groups, the dicotyledons and the monocotyledons. The origin of the angiosperms is as yet unknown. They are known to have existed in the Jurassic period, but were not at all abundant until the Cretaceous period (see **Paleobotany**). The earliest fossil members of this group are well-differentiated plants which give little indication as to their possible ancestry. Within the group, adaptation seems to be from the woody type to the herbaceous, and from plants with flowers having an indefinite number of parts arranged in spiral manner and not fused. As adaptation progressed the number of flower parts became reduced and definite and finally fused. In many cases great irregularity replaced the more primitive regularity. The angiosperms are the dominant land flora of the present day.

ANGLE (Mathematics). The figure obtained by drawing two straight lines, called the sides of the angle, from a point, called the vertex. In trigonometry, an angle measures the rotation of one straight line, the terminal line, about a fixed point on an initial line. It is positive if the direction of rotation is counterclockwise. A unit for measuring angles is the radian, which is that angle whose intercepted arc in a circle equals the radius of the circle. Thus π radians = 180°; 1 radian = $180°/\pi$ = $57.29578\cdots°$ or 57° 17′ 44.6″, approximately, and 1° is approximately 0.017453 radian.

If the magnitude of an angle equals 2π radians, it is called a perigon angle and such an angle, divided into 360 equal parts, has a magnitude of 360°. A right angle equals 90° or $\pi/2$ radians; a straight angle, 180° or π radians; an acute angle is less than 90°; an obtuse angle, greater than 90° (but frequently limited to one less than 180°). A reflex angle is greater than 180° but less than 360°. Oblique angle is a general term for one not equal to 90° or 180°. Related angles are designated as follows: conjugate, if their sum equals 360°; supplementary, if the sum equals 180°; complementary, if 90°; vertical, if they have a vertex in common and the sides of one angle are prolongations of the sides of the other.

An angle inside a circle is central, if its sides are radii and its vertex is at the center of the circle; inscribed, if its sides are chords and its vertex is on the circumference of the circle.

The previous definitions refer to a plane angle, which is usually meant by the word angle, without a qualifying adjective. However, there are several other kinds of angles as the subsequent discussion will show.

The angle between two intersecting planes is called a dihedral angle. The line of intersection of the planes is the edge of the angle and the planes are the faces of the angle. Two dihedral angles are adjacent if they have a common edge and face. If two planes are parallel, the dihedral angle between them is zero; if perpendicular, the two adjacent angles are right dihedral angles. The plane angle of a dihedral angle is formed by two straight lines, one in each plane, perpendicular to the edge at the same point. The terms vertical, acute, obtuse, complementary, supplementary, etc., as used for plane angles, are also applied to dihedral angles.

Analytically, the dihedral angle can be defined as

$$\cos\theta = \lambda\lambda' + \mu\mu' + \nu\nu'$$

where the direction cosines of perpendiculars to the two intersecting planes are (λ, μ, ν) and (λ', μ', ν'). In vector notation, the relation becomes

$$\cos\theta = \csc\phi_2 \csc\phi_3(\mathbf{e}_{12} \times \mathbf{e}_{23}) \cdot (\mathbf{e}_{23} \times \mathbf{e}_{34})$$

where $\mathbf{e}_{ij}$ is a unit vector drawn from point i to point j; $\mathbf{e}_{23}$ is on the edge of the dihedral angle; $\mathbf{e}_{12}$ on one face and $\mathbf{e}_{34}$ on the other; ϕ_2 is the plane angle determined by the first vector product and ϕ_3 that angle determined by the second vector product.

If three or more planes meet at a common point, a polyhedral angle exists. The vertex, edges, faces, and face angles are defined as for plane and dihedral angles. The polyhedral angle is convex if every section made by a plane cutting all of its edges is a convex polygon. It is called trihedral, tetrahedral, etc., when it has three, four faces, etc. A trihedral angle is rectangular, birectangular, trirectangular, if it has one, two, or three right dihedral angles. If the vertex of a trihedral angle is at the center of a sphere, its faces intersect the sphere in great circle arcs and spherical angles are formed (see **Triangle**).

Now consider a small cone with a base of area dS and a vertex at a fixed point P. The cone will cut out an area $d\sigma$ on a sphere of radius r with center at P. The angle subtended by dS at P is defined as $d\omega = d\sigma/r^2$ and is called a solid angle. It is numerically equal to the area cut out by the same cone on a sphere of unit radius at the same point P. The unit used for measuring a solid angle is the steradian. Since the area of a sphere of unit radius equals 4π, the total solid angle about a point is 4π steradians.

ANGLE OF DEPARTURE. The angle between the line of propagation of a radiowave and the earth's surface at the point of transmission.

ANGLERFISHES (*Osteichthyes*). All members of the anglerfish order (*Lophiformes*) have a modified single movable dorsal fin ray which carries a kind of "bait" at its end; this is the basis of the name *anglerfish.* The "bait" is known technically as the illicium. Other unusual modifications in this group include the pectoral fins, which enable the fishes to crawl along the ground. All anglerfishes are slow-moving, almost motionless fishes, attracting their prey by means of their natural bait, at which time they suck in the prey with an extremely fast motion. The mouth acts as a giant suction trap.

Three suborders are distinguished: (1) *Goosefishes*, with 1 family; (2) *frogfishes*, with 4 families; and (3) *deepsea anglerfishes*, with 10 families. Altogether there are 225 species of anglerfishes distributed for the most part in tropical, subtropical, and temperate waters of the world's seas.

It is generally believed that anglerfishes developed from perchlike fishes which took on the froglike shape as an adaptation to their life between rocks in coastal zones. According to this viewpoint, the modified pectoral fins act to ensure balance in the breaking waves. In further developments, the goosefishes adapted to living in deep coastal waters on the floor. The frogfishes adapted to living in dense plant growths, such as Sargassum seaweed; and the deepsea anglerfishes returned to a pelagic life, losing their dependence on a supporting object.

Goosefishes. These fishes are of suborder *Lophioidei*, family *Lophiidae.* They achieve a length exceeding 4 feet (1.2 meters) and a weight of about 45 pounds (20 kilograms), and are the largest anglerfishes. It has been reported that these fishes spend most of their lives lying on the bottom, where they blend into their surroundings so well that they can barely be perceived. Goosefishes are flattened from the back toward the belly (i.e., not laterally). The most striking part of the body is the head, the width of which is about two-thirds its length. The mouth opening extends over almost the entire width of the head; it has many needle-sharp teeth that seem capable of holding anything they grab. To attract prey, goosefishes use their bait, the illicium, which has been formed from the first six spiny rays of the dorsal fin. This consists of a line at the end of which dangles a fleshy, sometimes worm-shaped shred of skin, which is literally dangled in the water by the goosefish. If another fish approaches with the mistaken idea that this is a worm, the goosefish waits motionless until the prey is close enough to catch. This method of predation must be extremely successful, because goosefishes that are caught almost always have their stomachs full, so full in fact that their contents sometimes equal one third of the body weight of the fish. The prey include many species of fishes, crustaceans, cephalopods, and other organisms. When hungry, goosefishes gather in shallow water visited by diving birds,—because even small birds are eaten.

During the spawning period of the *allmouth* or *angler (Lophius piscatorius)*, which occurs in early summer along the European coasts and between January and February in the North Atlantic Ocean, the females develop a special kind of appetite. The fish leave their grounds near the coast and move into deeper regions (3280 and 6560 feet; 1000 and 2000 meters). The eggs are released in long, glandularly secreted tubes up to 1 foot (30 centimeters) wide, $\frac{1}{4}$-inch (5 millimeters) thick and as long as 13 feet (4 meters). The eggs are situated singly or by twos in six-sided compartments in a sort of honeycomb arrangement. Eventually, the structure tears and the eggs float individually in the water until the young hatch. The larvae, which are about $\frac{1}{8}$-inch (2 millimeters) long, swimming near the surface of the water do not resemble their parents at all; they look much more like typical fish larvae. After a four-month larval period, with growth up to a length of 4 to 6 inches (10 to 15 centimeters), and a complex change in body shape, the juvenile goosefishes have the adult appearance and behavior.

In spite of their unusual appearance, goosefishes are popular as food. The head and leathery skin are removed so that little remaining in the frying pans looks like the original fish. In Europe, the fishes are sold fresh or smoked as "trout sturgeon."

Frogfishes. These fishes are of the suborder *Antennarioidei* and are usually small, seldom exceeding a length of 12 inches (30 centimeters). All frogfishes have a peculiar body shape, enabling them to resemble their surroundings to such a high degree that they blend into the algae "forests" like the Sargasso frogfish. This mimicry of surroundings is created by many skin folds and various skin formations, supplemented by the ability to adapt to some extent in color and pattern in the background. Frogfishes which do not swim well are prevented from falling through the Sargasso seaweed or other objects by their pectoral fins, which function much like human hands in grabbing and holding objects. The reproduction and other life habits of this suborder remain essentially unknown. Female frogfishes observed in aquariums have produced spawning tubes similar to those in goosefishes.

Deepsea Anglerfishes. These fishes are of the suborder *Ceratioidei* and are characterized by the absence of pectoral fins and the fact that only females have a fishing organ. While goosefishes and frogfishes live primarily in shallow water and may migrate into deeper water only for spawning, the approximately 120 species of deepsea anglerfishes are found at substantial oceanic depths, ranging from 985 to 13,125 feet (300 to 4000 meters). Any typical bait on the end of a fishing line would no longer be recognizable at these depths, so deepsea anglerfishes have a luminous organ at the end of the line. Production of the light is not fully understood, but some authorities attribute this to the presence of luminous bacteria. See accompanying illustration.

Deepsea anglerfish (*Linophryne arborifer*).

Parasitic Males. The ceratiid anglers (*Ceratiidae*), photocorynid anglers (*Photocorynidae*), and linophrynid anglers (*Linophrynidae*) are species with very small males which parasitize the females and stay with them. The males do have their own gill respiratory system and the necessary vessels to supply their organs with oxygen, but their food is obtained from the bloodstream of their female hosts, with certain blood vessels in the males' heads in a dependent relationship with vessels in the females. Close examination of the dwarf males shows that several systems are more or less degenerated in them, particularly the teeth and the intestinal tract. They also lack the fishing line, as do all male deepsea anglerfishes.

The biological significance of parasitic males is related to the method of propagation of the species. Despite numerous signaling systems, including the luminous organs, it remains difficult for the different sexes to find each other in the deep sea, where all sunlight is absent. This dangerous disadvantage is offset by having just one partner seek food while the other parasitizes the food gatherer. Exemplary of the size differences is the female *Ceratias hollbolli*, which is some 40 inches (103 centimeters) long, as compared with the male, which averages about 3.5 inches (9 centimeters) in length. The size differences are not this large in some species.

The female deepsea anglerfishes are very active predators which feed on large organisms. Their jaws have powerful teeth and their stomachs are so greatly distensible that they can swallow prey which is larger than they are.

Deepsea anglerfishes primarily inhabit warmer parts of the Atlantic, Pacific, and Indian Oceans. Their numbers decrease sharply in the northern and southern temperate zones. Only a few species have been found in the north Atlantic Ocean.

Researches at the Department of Biology, California State University (Long Beach, California) studied the action of a species of Antennarius that originated in the Philippine waters. These investigators found that the pattern of movement of the illicial apparatus seems to be species-specific, ranging from simple strokes in the vertical plane to a complex triangular pattern, alternating with rapid sinusoidal thrusts. As reported by Pietsch and Grobecker (1978), "During a single luring sequence the illicium is initially brought straight forward in front of the mouth of the angler, and the bait is rapidly vibrated for 1 or 2 seconds. The bait is then held nearly motionless as the illicium is slowly laid back again onto the head and returned to its nonluring position. When the animal is sufficiently aroused, however, the bait makes a large and rapid sweeping motion that describes a nearly perfect circle. The thin membranous quality of the bait allows it to ripple while being pulled through the water, simulating the lateral undulations of a swimming fish. The lure thus provides not only a highly attractive visual cue but presumably also a low-frequency pressure stimulus for potential prey." This is an example of great energy savings in the quest for food. Further details are given in *Science*, **201**, 369–370 (1978).

In 1979, researchers at the College of Fisheries, University of Washington (Seattle, Washington), made a high-speed photographic study of the feeding actions of antennariid anglerfishes. Light cinematography at 800 and 1000 frames per second was used. It was found that single feeding events for the anglerfishes occur at speeds greater than four times those described for other fishes. Further details are given in *Science*, **205**, 1161–1162 (1979).

ANGLESITE. Naturally occurring lead sulfate ($PbSO_4$), which crystallizes in the orthorhombic system and may be found mixed with galena, from which it is usually formed by oxidation. Hardness, 3; specific gravity, 6.12–6.39; luster, adamantine to vitreous or resinous; transparent to opaque; streak, white; colorless to white or green, but rarely may be yellow or blue. This mineral is used as a source of lead.

Anglesite, whose name derives from Anglesey, England, is found in many European localities; in the United States it has been found in large crystals in the Wheatley Mine, Phoenixville, Pennsylvania, and also in Missouri, Utah, Arizona, and Idaho.

ANGLE (Slip). 1. The angle included between the direction of the applied force and the surface of shear during the plastic flow of a solid body. 2. The angle of repose. See also **Repose (Angle of).**

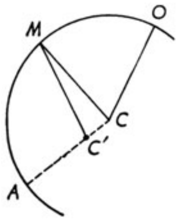
ANGLE-WING (*Insecta, Lepidoptera*). Butterflies of the genus *Polygonia*. Their wings are sharply angular but no more so than those of some other species.

ANGULAR DISTRIBUTION (Particle). In physics, the distribution in angle, relative to an experimentally specified direction, of the intensity of particles or photons resulting from a nuclear or an extranuclear process. Commonly, the specified direction is that of an incident beam, and the angular distribution is that of particles which are scattered or are the products of nuclear reactions. Alternatively, the specified direction might be that of an applied field, or a direction of polarization, or the direction of emission of an associated radiation.

ANGULAR MAGNIFICATION. The ratio of the tangent of the angle with the optical axis made by a ray upon emergence from an optical instrument to the tangent of the angle for the conjugate incident ray.

ANGULAR MEASUREMENT (Eccentricity Correction). A correction for eccentricity must be applied to many types of instruments used for astronomical angular measurement. Instruments for this purpose usually consist of a circle graduated in angular units, and an arm, assumed to be concentric with the circle, that sweeps around the circle, carrying a vernier, or a measuring microscope, for the purpose of determining accurately the direction of the arm relative to the circle. It is practically a mechanical impossibility to make the centers of the circle and the measuring arm exactly coincident.

In the diagram, point *C* is the center of the circle *OMA*, graduated from *O*. *C′* is the center of the measuring arm (commonly known as the alidade). The direction *C′M* is the actual direction of the alidade, and *OM* is the direction obtained from the circle reading (i.e., the angle *OCM*). The difference between these two directions, the angle *C′MC*, is the eccentricity correction for the circle reading *OM*. This will be different for different circle readings, being zero for the circle reading *OA*.



Eccentricity correction for circle with alidade.

An equation may be developed that will give the eccentricity correction for any circle reading as a function of the reading and three numerical constants. To determine these constants, at least three known angles must be measured with the instrument. The differences between the values obtained with the instrument and the known values of the angles are the eccentricity corrections for the circle readings. These eccentricity corrections are then used for the solution of three equations for the three constants. With the constants determined, the equation, giving the eccentricity correction for any circle reading, may be written. The results are usually tabulated, or plotted on a curve and supplied by the maker of the instrument.

ANGULAR MOMENTUM (Particle). The vector product of the instantaneous values of the position vector and the linear momentum,

$$\mathbf{M} = m\mathbf{r} \times \mathbf{v}$$

For an interacting system of particles, the law of the conservation of angular momentum states that the rate of change of the total angular momentum equals the vector sum of the moments of the external forces applied to the system,

$$\frac{d}{dt}\sum_{s} M_s = \sum v_s \times F_s$$

where the summation is over the particles composing the system. In the absence of external forces, the angular momentum remains constant and no change of configuration can alter the total angular momentum of the system. Thus, a slowly rotating swarm of particles, like a cloud of gas in space, if it contracts under its own gravitational attraction, must rotate the more rapidly to keep its angular momentum constant. In the atmosphere, rotating air-masses tend to preserve their absolute angular momentum and the strong winds in hurricanes arise from the convergence of air into the lower levels.

If the particles are bound together in a rigid body, it is convenient to define the moment-of-inertia tensor

$$I_{ij} = \sum_{s} m_s \left(\delta_{ij} v_e v_e = v_i v_j\right)_s$$

where the position vectors are relative to the center of mass. Then the total angular momentum due to motion about the center of mass is

$$M_i = T_{ij}\omega_j$$

where ω_j is the instantaneous angular velocity of the rigid body (repeated tensor suffixes indicate summation).

The principle of conservation applies to angular as well as to linear momentum. That is, no change of configuration within a system, uninfluenced by external forces, can alter the total angular momentum of the system. Thus, a slowly rotating swarm of particles, like a cloud of gas in space, if it contracts under its own gravitational attraction with attendant decrease in moment of inertia, must rotate the more rapidly to keep its angular momentum constant. Again, if a person, whirling about on tiptoe, with arms extended, suddenly brings the arms down to the sides, the individual will as suddenly begin to whirl faster, the effect being more pronounced if heavy weights are held in the hands. Angular momentum being a vector quantity, the principle applies as well to its direction as to its magnitude. The result is that any rotating body tends to maintain the same axis of rotation, a fact well illustrated by the spinning top and by the stabilizers used on some ocean vessels.

For an elementary or other particle, angular momentum may arise from (a) rotation about an axis, (b) revolution in an orbit, or from both (a) and (b). The angular momentum of rotation is called intrinsic angular momentum, or spin. When a nucleus is considered as a single particle, its total angular momentum is also referred to as its spin. For an elementary particle, the component in a particular direction of both kinds of angular momentum is quantized; the quantum of spin angular momentum is $\frac{1}{2}$, and the quantum of orbital angular momentum is .

ANGULAR VELOCITY AND ANGULAR ACCELERATION. Quantities relating to rotational motion. While the use of the term "angular velocity" may be extended to any motion of a point with respect to any axis, it is commonly applied to cases of rotation. It is then the vector, whose magnitude is the time rate of change of the angle θ rotated through, i.e., $d\theta/dt$, and whose direction is arbitrarily defined as that direction of the rotation axis for which the rotation is clockwise. The usual symbol is $\boldsymbol{\omega}$ or $\boldsymbol{\Omega}$.

The concept of angular velocity is most useful in the case of rigid body motion. If a rigid body rotates about a fixed axis and the position vector of any point *P* with respect to any point on the axis as origin is $\mathbf{r}$, the velocity $\mathbf{v}$ of *P* relative to this origin is $\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$, where $\boldsymbol{\omega}$ is the instantaneous vector angular velocity. This indeed may serve as a definition of $\boldsymbol{\omega}$.

The average angular velocity may be defined as the ratio of the angular displacement divided by the time. In general, however, this is not a vector, since a finite angular displacement is not a vector. The instantaneous angular velocity is more widely used.

Angular velocities, like linear velocities, are vectorially added; for example, if a top is spinning about an axis which is simultaneously being tipped over toward the table, the resultant angular velocity is the vector sum of the angular velocities of spin and of tipping. (This enters into the theory of precession.)

Angular acceleration is the time rate of change of the angular velocity, expressed by the vector derivative $d\boldsymbol{\omega}/dt$. Only in case the direction of the axis remains unchanged can the angular velocity and angular acceleration be treated as scalars. The effect of torque applied to a body free to rotate about an axis is to give it angular acceleration, and the opposition offered by the body to this process gives rise to the concept of moment of inertia. See also **Stroboscope;** and **Velocity and Speed Measurement.**

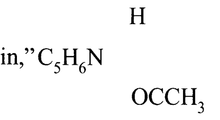
ANHEDRAL. Minerals in igneous rocks which are not bounded by their typical crystal faces. Such minerals are said to be anhedral or *allotriomophic*, the latter term (now obsolete) proposed by Rosenbusch in 1887.

ANHYDRITE. The mineral anhydrous calcium sulfate, $CaSO_4$, occurs in granular, scaly, or fibrous masses, is rarely crystallized in orthorhombic tabular or prismatic forms. Hardness, 3–3.5; sp gr, 2.9–2.98; translucent to opaque; streak white; color, white, gray, bluish, or reddish. Anhydrite has three cleavages at right angles to one another. It is similar to gypsum and occurs under the same conditions, often with the latter mineral. It is usually found in sedimentary rocks associated with limestones, salt, and gypsum, into which it changes slowly by the absorption of water. See also **Gypsum.**

Anhydrite is found in Poland, Saxony, Bavaria, Württemberg, Switzerland, and France; in the United States, in South Dakota, New Mexico, Texas, New Jersey, and Massachusetts; in Canada, in Nova Scotia, New Brunswick, and exceptional specimens from the Faraday Uranium Mine near Bancroft, Ontario.

ANILINE. Aniline, phenylamine, aminobenzene, $C_6H_5NH_2$, is a colorless, odorous liquid, an amine, with melting point $-6°C$, boiling point 184°C, is slightly soluble in water, miscible in all proportions with alcohol or ether, poisonous, turns yellow to brown in the air, is a weak base forming salts with acids, e.g., anilinehydrochloride ("aniline salt," $C_6H_5NH_2 \cdot HCl$) from which aniline is reformed by addition of sodium hydroxide solution. Aniline reacts (1) with hypochlorite solution, to form a transient violet coloration, (2) with nitrous acid (a) warm, to form nitrogen gas plus phenol, (b) cold, to form diazonium salt (benzene diazonium chloride, $C_6H_5N—Cl$), (3) with acetyl chloride, acetic anhydride, or acetic acid glacial, to form N-phenylacetamide

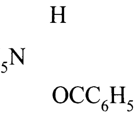
acetanilide, "antifebrin,"



$$C_5H_6N\begin{array}{l}H\\ \\OCCH_3\end{array}$$

(4) with benzoyl chloride, to form N-phenylbenzamide

benzanilide,



$$C_5H_5N\begin{array}{l}H\\ \\OCC_6H_5\end{array}$$

(5) with benzenesulfonyl chloride, to form N-phenylbenzene sulfonamide, $C_6H_5SO_2NHC_6H_5$, soluble in sodium hydroxide, (6) with chloroform, $CHCl_3$, plus alcohol plus sodium hydroxide, to form phenyl isocyanide, C_6H_5NC, very poisonous, (7) with H_2SO_4 at 180° to 200°C, to form para-aminobenzene sulfonic acid (sulfanilic acid, $H_2N \cdot C_6H_4 \cdot SO_2H$ (1,4)), (8) with HNO_3, when the amine group is protected, e.g., using acetanilide, to form mainly paranitroacetanilide, $CH_3CONH \cdot C_6H_4 \cdot NO_2$ (1.4), from which paranitroaniline, $H_2N \cdot C_6H_4 \cdot NO_2(1,4)$ is obtained by boiling with concentrated hydrochloric acid, (9) with chlorine in an anhydrous solvent, such as chloroform or acetic acid glacial, to form 2,4,6-trichloroaniline $(1)H_2N \cdot C_6H_2Cl_3(2,4,6)$, (10) with bromine water, to form white solid 2,4,6-tribromoaniline, $(1)H_2N \cdot C_6H_2Br_3(2,4,6)$, (11) with potassium dichromate in sulfuric acid, to form aniline black dye, and, by further oxidation, benzoquinone, $O{:}C_6H_4{:}O(1,4)$, (12) with potassium permanganate in sodium hydroxide, to form azobenzene, $C_6H_5N{:}NC_6H_5$, along with some azoxybenzene $C_6H_5NO{:}NC_6H_5$, (13) with reducing agents, to form aminohexahydrobenzene (cyclohexylamine, $H_2N \cdot C_6H_{11}$), (14) with alkyl halides or alcohols heated, to form alkyl anilines, e.g., methylaniline, $C_6H_5NHCH_3$, dimethylaniline, $C_6H_5N(CH_3)_2$.

Aniline may be made (1) by the reduction, with iron or tin in HCl, of nitrobenzene, and (2) by the amination of chlorobenzene by heating with ammonia to a high temperature corresponding to a pressure of over 200 atmospheres in the presence of a catalyst (a mixture of cuprous chloride and oxide). Aniline is the end-point of reduction of most mono-nitrogen substituted benzene nuclei, as nitrosobenzene, beta-phenylhydroxylamine, azoxybenzene, azobenzene, hydrazobenzene. Aniline is detected by the violet coloration produced by a small amount of sodium hypochlorite.

Aniline is used (1) as a solvent, (2) in the preparation of compounds as illustrated above, (3) in the manufacture of dyes and their intermediates, (4) in the manufacture of medicinal chemicals. See also **Amines.**

ANIMALCULE. A minute animal. Applied to the protozoa and to such microscopic forms as the rotifers. See also **Rotatoria.**

Bear animalcule.

ANIMAL UNIT. A term sometimes used in pasture and forage land management. One animal unit equals one mature cow, or one horse, or five sheep, or two yearling calves.

ANIMAL-UNIT-MONTH. The feed or forage needed to support one animal unit (see **Animal Unit**) for 30 days. One animal-unit-month is roughly equivalent to 0.3 ton (0.27 metric ton) of hay, or 300 pounds (135 kilograms) of total digestible nutrients.

ANION. A negatively charged atom or radical. In electrolysis, an anion is the ion which deposits on the anode; that portion of an electrolyte which carries the negative charge and travels against the conventional direction of the electric current in a cell. Within the category of anions are included the nonmetallic ions and the acid radicals, as well as the hydroxyl ion, OH^-. In electrochemical reactions, they are designated by the minus sign placed above and after the symbol, such as Cl^- and SO_4^{2-}, the number of the minus sign indicating the magnitude, in electrons, of the electrical charge carried by the anion. In a battery, it is the deposition of negative anions that makes the anode negative. See also **Ion.**

ANION-EXCHANGE RESINS. See **Ion Exchange Resins.**

ANISE. Of the family *Umbelliferae* (carrot family), the anise plant (*Pimpinella anisum*) is native to the Mediterranean region and is cultivated in Egypt, Malta, Spain, and Syria, but also in other areas of the world, such as Germany and the United States. (This plant should not be confused with *fennel* or *finocchio*, which is commonly called anise in the marketplaces of the United States.)

Frequently, the plant is grown by gardeners of small plots as part of an herb garden. The plant achieves a height of about 2 feet (0.6 meter) with several slender branches. See accompanying illustration. The aromatic, warm, and sweetish odor and taste of the seed, leaves, and stem arises from the presence of a volatile oil that contains anethole (*p*-propenyl phenylmethyl ether, $C_3H_5C_6H_4OCH_3$), the derivatives of which (anisole and anisaldehyde) are used in food flavoring, particularly bakery, liqueur, and candy products, as well as ingredients for perfumes. For commercial production of anise oil, the seeds and the dried, ripe fruit of the plant are used. Anise oil, a colorless to pale-yellow, strongly refractive liquid of characteristic odor and taste, is prepared by

Anise plant (*Pimpinella anisum*).

steam distillation of the seed and fruit. The oil contains choline which finds use in medicine as a carminative and expectorant.

The anise plant is an annual, planted directly from seed in the spring. The leaves of the plant can be used directly in salads to provide a distinctive flavor.

The fruit of a small evergreen tree (*Illicium anisatum*) of the Magnolia family is the source of *star* or *Chinese anise*. The aromatic and chemical characteristics of this plant are similar to those of *Pimpinella anisum* and thus there are similar uses for it.

ANISODESMIC STRUCTURE. A type of ionic crystal in which some of the ions tend to form tightly bound groups, e.g., nitrate and chlorate.

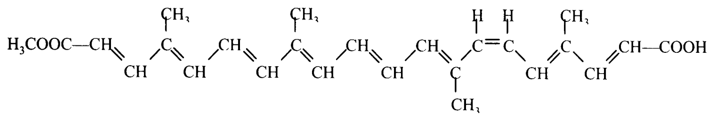
ANISOTROPIC MEDIUM. An anisotropic medium has different optical or other physical properties in different directions. Wood and calcite crystals are anisotropic, while fully-annealed glass and, in general, fluids at rest are isotropic.

ANNABERGITE. The mineral annabergite is a rather rare nickel arsenate with the formula $Ni_3(AsO_4)_2 \cdot 8H_2O$, crystallizing in the monoclinic system. It is of secondary origin, resulting from the alteration of pre-existing nickel minerals, commonly found as surface alteration crust on nickeline. Annabergite has been found in Saxony, France, including Annaberg, from which its name is derived, and as exceptional crystals at Laurium, Greece, and in Cobalt, Ontario, Canada.

ANNATTO FOOD COLORS. These colors are natural carotenoid colorants derived from the seed of the tropical annatto tree (*Bixa orellana*). The surface of the seeds contains a highly colored resin, consisting primarily of the carotenoid *bixin*. The bixin is extracted from the seed by a special process to produce a pure, soluble colorant. Bixin, one of the relatively few naturally occurring *cis* compounds, has a chemical structure similar to the nucleus of carotene with a free and esterified carboxyl group as end groups. Its formula is $C_{25}H_{30}O_4$.

Bixin is an oil-soluble, highly stable coloring ingredient. The saponification of the methyl ester group to form the dicarboxylic acid yields the water-soluble form of bixin, sometimes called *norbixin*. Annatto colorants date back into antiquity. The colorant has been used for centuries in connection with various textiles, medicinals, cosmetics, and foods. Annatto colors have also been used to color cheese, butter, and other dairy products for over a century. See also **Carotenoids.**

Processors make annatto colors available as a refined powder, soluble in water at pH values above 4.0 (solubility about 10 grams in 100 milliliters of distilled water at 25°C), in an acid-soluble form, in an oil-soluble form, in a water- and oil-soluble form, and in a variety of hues ranging from delicate yellows to hearty orange. Annatto extract is frequently mixed with turmeric extract to obtain various hues.



Structure of bixin.

ANNEALING. The process of holding a solid material at an elevated temperature for a specified length of time in order that any metastable condition, such as frozen-in stains, dislocations, and vacancies may go into thermodynamic equilibrium. This may result in recrystallization and polygonization of cold-worked materials.

Annealing generally falls into the technology of heat treatment and varies with materials and the intended end uses of the materials, as well as the prior processing of them. In the case of nonferrous alloys, annealing is primarily a heat treatment for the purpose of removing the hardening due to cold work. Annealing also may be used with nonferrous precipitation hardening alloys to cause softening through agglomeration of the hardening constituent into fewer and larger particles.

Ferrous Metallurgy. In the case of ferrous materials, the term annealing usually implies full annealing. This heat treatment involves a change of phase inasmuch as the metal is heated into the austenitic region. Cooling slowly back to room temperature then develops a softened structure of pearlite and ferrite. The annealing of cold-worked metal is termed *process annealing*, wherein a change of phase is not involved. Annealing takes several forms in terms of the time-temperature relationships imposed upon the materials. *Box annealing, isothermal annealing, normalizing, patenting, spheroidize annealing*, and *stress relieving* are described under **Iron Metals, Alloys, and Steels.**

Annealing of Cold-Worked Metals. Ductile metals hardened by cold-working may be softened by annealing. Annealing is often an important intermediate step in producing metals by cold deformation. Thus, in the formation of fine wires through wire drawing, several intermediate anneals may be required. Annealing may also be the last production step when metal objects are desired in a final softened condition.

In general, cold working increases manyfold the dislocation density of a metal. A severely cold-worked metal may easily have a dislocation density 10^6 times greater than in the same unworked metal. Since each dislocation is surrounded by a strain field extending over long distances on an atomic scale, each dislocation contributes to the strain energy of the metal and, accordingly, to its free energy. When the metal is annealed, the free energy associated with the dislocations resulting from cold work furnishes a driving force that can effectively reduce the dislocation density back to the value that existed before deformation.

Three basic stages are generally recognized as occurring during the annealing of cold worked metals. These are recovery, recrystallization, and grain growth.

In recovery, the strain energy is lowered by the recombination of dislocations of opposite sign, or by rearrangements of dislocations into configurations of lower strain energy. A simple well-known example of this latter is the polygonization of the dislocations in a bent crystal. When a crystal is bent, the curved shape is the result of the accumulation, upon the slip planes of the crystal, of a large number of edge dislocations of the same sign. During recovery, these dislocations move from their more or less random positions along the slip planes into a set of vertical walls normal to the slip planes. This movement is accomplished by both slip and dislocation climb. The walls of dislocations that are formed in this manner constitute a form of grain boundary across which the crystal lattice is slightly rotated by the order of minutes of arc. Such boundaries are better known as subgrain boundaries. The crystalline material between these subboundaries is effectively free of dislocations. It is thus apparent that polygonization transforms a highly strained bent crystal into a set of small subgrains that are nearly strain-free. (See Fig. 1.)

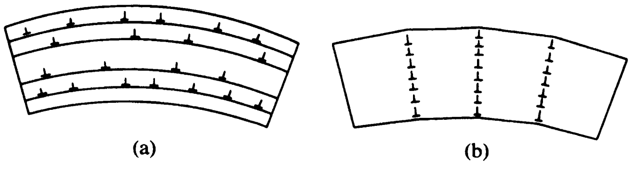


Fig. 1. Realignment of edge dislocations during polygonization: (a) excess edge dislocations that remain on active slip planes after a crystal is bent: (b) arrangement of dislocations after polygonization.

The rate of recovery is normally highest at the start of an isothermal annealing cycle because the driving force is largest at that time. As recovery continues, the driving force diminishes as the available strain energy is used up and the rate of recovery falls continuously toward zero. A plot of the rate of recovery as a function of time yields a curve that is somewhat similar in appearance to an exponential decay curve. The rate of recovery is also temperature dependent and may be expressed, in a number of cases, by a simple empirical equation of the form

$$1/t = Ae^{-Q/RT}$$

where t is the time to attain a certain fixed amount of recovery, A is a constant, R, the universal gas constant, T, the absolute temperature, and Q, an empirical activation energy. Because the reactions that occur during recovery are complex, it is usually not possible to attach a simple meaning to the activation energy for recovery.

Recrystallization is the process whereby the distorted grains or crystals of a cold-worked metal are reconverted into new (essentially) strain-free grains. It occurs by the nucleation of minute submicroscopic crystals that grow out into and consume the strained material surrounding them. Recrystallization is, therefore, a nucleation and growth phenomenon and, characteristically, the rate of recrystallization starts slowly, builds up to a maximum, and then diminishes back to zero. The increase in the rate at the early stages of recrystallization is due primarily to continued nucleation of new grains while the older ones continue to grow. The final falling off in the rate is the result of the progressive consumption of the material available for recrystallization. A metal is said to be completely recrystallized when all of the original deformed structure has been eliminated.

Recrystallization, like recovery, is thermally activated and occurs at a rate that grows very rapidly with increasing temperature, as may be seen in the accompanying diagram where the amount of recrystallization in copper is plotted as a function of the time for six different temperatures. (See Fig. 2.)

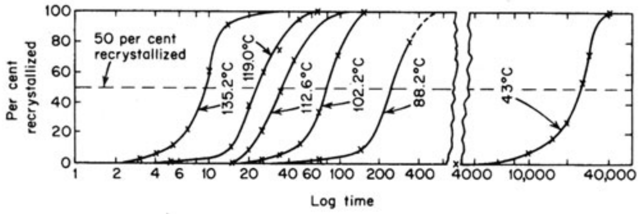


Fig. 2. Complex polygonized structure in a single silicon-iron crystal deformed 8% by cold rolling before being annealed 1 hour at 1100°C.

The driving force for recrystallization also comes from the strain energy of the excess dislocations created by cold work. It is therefore apparent that recovery and recrystallization are competitive processes. Usually a metal may undergo a considerable degree of recovery before visible evidence of recrystallization is obtained. However, since the recrystallized grains grow from very small beginnings, the recrystallization process undoubtedly is occurring long before it can be detected visually. Also, there is reason to believe that the nuclei of the recrystallized grains may be formed as a result, at least in some cases, of processes related to recovery. It should also be noted that recovery phenomena may continue to occur during recrystallization in those grains not yet consumed by the recrystallization process. The degree to which the manifestations of recrystallization and recovery appear to overlap is a function of the metal concerned and of the nature of the deformation that it has received. Under certain conditions, it is possible to have recovery occur without recrystallization. This is particularly true when the amount of deformation is insufficient to cause recrystallization, or when the type of deformation, although extensive, is very simple, as in the case of a zinc or magnesium crystal deformed only by slip on the basal plane. Examples have been observed where single crystals of these metals have been deformed in this manner by as much as 700% and still failed to recrystallize on annealing.

After a metal has undergone recrystallization it can still undergo grain growth. The driving force in this case comes from the surface energy of the grain boundaries. A close analogy exists between the growth of grains in a metal during annealing and the growth of soap bubbles in a soap froth. In the soap froth, a small bubble that finds itself surrounded by larger neighbors will normally have but a few sides convex toward the bubble center. This curvature produces a small but finite excess gas pressure inside the small bubble, which causes gas to diffuse through the bubble wall into the neighboring larger bubbles. The smaller bubble, consequently, grows smaller and disappears, while the larger ones surrounding it grow in size. At the same time, the average bubble in the froth must also grow in size. The same basic phenomenon occurs during grain growth in metals. In this case, the atoms from the smaller grains move across the grain boundaries and become part of the crystals of the larger grains. At all times a geometrically similar distribution of grain sizes exists in the metal, ranging from small to large, which promotes continued grain growth. However, as the average size increases, there is a corresponding decrease in the growth rate. This is easily understood in terms of the soap froth analogy because with an increase in bubble size there is a corresponding decrease in the average bubble wall curvature and in the pressure difference across bubble walls. It may also be shown that in a soap froth the average bubble size should increase as the square root of the time. This one-half power law, however, is seldom observed during grain growth in a metal. Usually the grain growth exponent (power to which the time is raised) is much smaller than one half, signifying that the empirical growth rates are much lower than would be expected from the soap froth analogy. Several reasons may be proposed in explanation of this fact. The motion of grain boundaries is known to be easily influenced by both impurity atoms in solid solution or present as small intermetallic inclusions. In either case, the grain boundary mobility is lowered with a corresponding decrease in the value of the grain growth exponent.

Under the proper conditions, a limiting grain size may be attained in a metal at which point grain growth ceases. This is often true in very thin specimens when the average grain diameter approaches the thickness of the specimen. At this time, the grain boundary geometry becomes two- instead of three-dimensional which reduces the average curvature and hinders further growth. Alternatively, it is possible for grain boundaries to become so held up by the nonmetallic inclusions that further growth is prevented. This condition is only achieved after a critical grain size has been achieved.

Associated with the limiting grain size effect mentioned above is a phenomenon known as secondary recrystallization. Sometimes, after a limiting grain size has been attained, a few grains may begin to grow again and may obtain very large sizes. This is actually not a true recrystallization but rather an unusual manifestation of grain growth. This formation of a new set of very large grains in material where growth had apparently ceased is known as secondary recrystallization.

Annealing of Glass. As with metals, glass is fabricated at high temperatures and is annealed to relieve stresses which would develop if the glass were permitted to cool in an uncontrolled fashion. If not annealed, products made from high-expansion glasses can break spontaneously as they cool freely in air. In annealing glasses, they are raised to an annealing point temperature and then cooled gradually to a temperature that is somewhat below the strain point. Usually, the rate of cooling within this range determines the magnitude of residual stresses after the glass arrives at room temperature. Once below the strain point, the cooling rate is limited only by any transient stresses that may develop. A typical time-temperature glass-annealing curve is shown in Fig. 3. Nor-

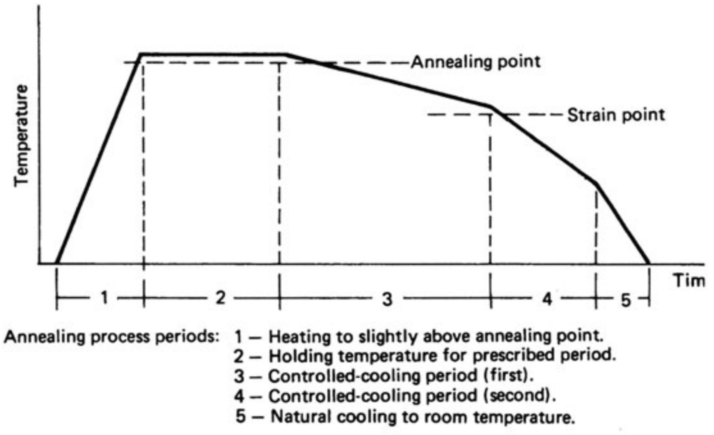


Fig. 3. Typical glass annealing curve.

mally, glass for optical purposes is annealed much more slowly than commercial glassware to improve the optical homogeneity of the material. In the annealing process, glass normally is heated and held at a temperature slightly higher than the annealing point temperature and controlled cooling is effected to a temperature slightly below the strain point to accommodate for differences in materials and as an extra safeguard. Controlled cooling may occur over two periods as indicated by the diagram.

Additional Reading

Banerjee, B. R.: "Annealing Heat Treatments," *Metal Progress*, **118**, 6, 58–64 (1980).

Bardes, B. P., Editor: "Heat Treatment of Carbon and Alloy Steels," in *Metals Handbook*, 9th edition, American Society for Metals, Metals Park, Ohio, 1979.

Chandler, H. E.: "Heat Treating Buyers Guide and Director," American Society for Metals, Metals Park, Ohio, 1981.

Chandler, H. E.: "Harness Retained Heat to Save Fuel and Time in Annealing Forgings," *Metal Progress*, 47–49 (February 1986).

McGannon, H. E.: "The Making, Shaping and Treating of Steel," 9th edition, U.S. Steel Corporation, Pittsburgh, 1971.

Staff: "Heat Treatment '79," American Society for Metals, Metals Park, Ohio, 1979.

Staff: "Heat Treating, Cleaning and Finishing," Vol. 2 of *Metals Handbook*, 8th edition, American Society for Metals, Metals Park, Ohio, 1981.

Staff: "Trends in Heat Processing Technology," *Metal Progress*, **119**, **1**, 102–106 (1981).

Wilson, R.: "Metallurgy and Heat Treatment of Tool Steels," McGraw-Hill, New York, 1975.

ANNELIDA. The segmented worms, including earthworms and leeches. This phylum is biologically interesting because it shows in a primitive form the structural plan of the more complex animals.

1. Individual of *Autolytus* with male about to detach (*Verrill, "Invertebrate Animals of Vineyard Sound".*) 2. Tufted worm (*Amphitrite ornata*). (*Drawn by Verrill.*)

The annelids are characterized by: (1) metameric segmentation; (2) a closed tubular circulatory system in most forms; (3) a coelom; (4) an excretory system with tubules opening from the coelom to the exterior in various segments; (5) a tubular alimentary tract, with regions specialized for various functions; (6) a nervous system consisting of a dorsal brain above the esophagus, connected by cords passing around the esophagus, with a ventral chain of ganglia (see **Ganglion**) below the alimentary tract; (7) setae present in many species.

The annelids are classified as follows:

Class *Archiannelida*. Small marine annelids without setae; few to many segments.

Class *Polychaeta*. Worms with setae. No suckers. External segmentation distinct and metameric. Earthworms and many aquatic species.

Class *Oligochaeta*. Earthworms and many aquatic species.

Class *Hirudinea*. Flattened worms without setae but with a sucker at each end of the body. External segmentation consisting of 2–14 annuli to each metamere. Mostly aquatic, a few marine and a few terrestrial. Mostly blood-sucking parasites. The leeches.

ANNIHILATION. A term used in physics to describe a process in which a particle and antiparticle combine and release the energy associated with their rest masses. The most common example is the annihilation of an electron pair. Usually the negatron and the positron of the pair first form an atom of positronium from which state they merge and are annihilated. To conserve both energy and momentum the rest mass of this particle and antiparticle is converted into two photons moving in opposite directions, each with an energy of 0.511 MeV. The energy associated with the annihilation of other particle-antiparticle pairs, such as a proton and antiproton, is much larger than for an electron pair and is carried away by pions or kaons. See also **Particles (Subatomic).**

ANNONA. Genus of the family *Annonaceae* (custard-apple family). This genus contains shrubs and small trees, many of which bear fruits that are consumed by humans. Generally, they are of relatively minor importance commercially, although some species are valued in regional markets and some are used to make purees which find use in a variety of processed foods. These fruits are composed of many individual ovaries which are more or less sunk in the fleshy receptacle and united to it and to each other. In some species, these collective fruits are quite large—up to 8 inches (20 centimeters) in length. Some of the fruits are so heavy that they drag down the branches. In tropical areas where they are grown, the small fruit trees frequently attract ants which must be destroyed, particularly during the fruiting season. The trees are susceptible to various fungus diseases, causing fruit rot. Copper-based fungicides are frequently used. *Annona* species are found in the tropics of both the northern and southern hemispheres.

Soursop. This species, *Annona muricata*, is found on the islands of the Caribbean Sea, in Florida, and in southeastern Asia, notably in Malaysia. The *A. muricata* is a small evergreen tree about the size of a peach tree. The leaves are leathery and malodorous. The fruit is large (8 inches; 20 centimeters in length) heavy, and pear-shaped, with a rough skin that has spinelike projections. The flesh is white, succulent, acidic, and is variously reported (depending upon particular variety) as having a rich flavor of wine, a taste reminiscent of the black currant, and a flavor something like that of a mango. The seeds are large and of a dark color. The tree has many branches, is decorative, and is valued as a garden ornament as well as for its fruit. In the United States, successful cultivation is confined essentially to southern Florida. The fruit can be found in the markets in the vicinity of Key West. The soursop is processed to remove seeds and fibers in preparation of a puree. This puree can be preserved by canning or freezing. It is then available for use as a base material for flavoring sherbets, ice creams, beverages, and other products. Prior to freezing, the puree is heated to a temperature of about 185°F (85°C) for a few minutes to inactivate peroxidase which, if present, may cause a pink discoloration and off-flavors. Sometimes soursop nectar is available in cans or bottles. The soursop is recommended by specialists at the Food and Agriculture Organization (United Nations) for inclusion in school and demonstration gardens in various developing tropical areas, such as found in west Africa.

Common Custard Apple. This species, *Annona reticulata*, is a deciduous tree that reaches a height of from 15 to 25 feet (4.5 to 7.5 meters). The tree is one of the more robust of the *Annona* species and is commonly found growing in the West Indies and many tropical and subtropical areas of the world. There are numerous trees of this species in southern Florida. The *A. reticulata* is an exception among the *Annona* species in that it spreads spontaneously and need not be individually planted. This widespread growth is exemplified by the profusion of the trees found in the forests of the Philippines, on the island of Guam, and in Malaysia. The fruit or custard apple is inferior in flavor and other edible characteristics as compared with the sugar apple and the cherimoya. The fruit is from 3 to 5 inches (7.5 to 12.5 centimeters) in diameter, has a smooth skin that is geometrically divided into rhomboid or hexagonal areoles. The color varies from red to reddish-brown when ripe. The pulp is semisweet with a tallowlike consistency. The seeds adhere tenaciously to the flesh.

Sweetsop or Sugar Apple. This species, *Annona squamosa*, is native to tropical America and does not do well in subtropical regions as will some of the *Annona* species. The plant is a small deciduous tree which attains a height of from 15 to 20 feet (4.5 to 6 meters). In addition to the West Indies, this species is found throughout the tropics in southern

Asia and is particularly popular in India where locally it is called *custard apple*, a fact that tends to confuse it with the true custard apple previously described. The fruit is regarded more highly than *A. reticulata*. The sweetsop is much smaller than the soursop fruit, ranging from 2 to 3 inches (5 to 7.5 centimeters) in diameter and has an almost smooth, segmented skin. The flesh is fragrant and sweet to the taste. The pulp is a pale-yellow and looks very much like custard. Seeds are dark-brown. An advantage of this species is that it produces fruit throughout the year.

Cherimoya. This species, *Annona tripétala* Ait., is a larger tree, rising to a height of about 25 feet (7.5 meters) and more. It is commonly found in the Peruvian Andes and surrounding regions. The plant is also planted in Mexico, a number of Central American countries, Hawaii, India, the Canary Islands, and on the Island of Madeira. The cherimoyas of Madeira are well known and valued for their elegant flavor. In Madeira, the cherimoya plant is trained on trellises much as grapes are grown. The fruit is variously shaped—sometimes spheroidal, ovoid, conoid, and heart-shaped. A sometimes used synonym for the fruit is *bullocksheart*. The fruit has a slightly mottled, comparatively smooth surface. The pulp is white with a pleasing acidulous taste. The seeds are easily separated from the pulp, an advantage in eating the fruit fresh as well as for processing. Over the last several decades, considerable success has been enjoyed in raising this species in California.

Other varieties of *Annona* which bear edible fruit of varying quality and desirability and of very minor importance include *A. montana*, *A. purpurea*, *A. glabra* (alligator-apple), *A. diversifolia*, *A. longifolia*, among several others.

Suggested Reading

Bailey, L. H.: "Standard Cyclopedia of Horticulture," Macmillan, New York, 1963.

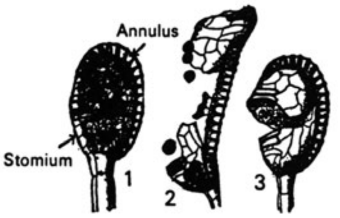
Little and Wadsworth: "Common Trees of Puerto Rico and the Virgin Islands," Agriculture Handbook 249, U.S. Forest Service, Washington, D.C., 1964.

ANNUAL. A plant which normally completes its life cycle, from seed to seed, in a single growing season. Typical annuals are corn (maize), wheat, cucumber, and nasturtium. Annual plants are especially suited for life in regions where the growing season is short and alternates with an unfavorable cold period or dry season.

ANNUAL RING. A layer of wood added to the stem in one growing season.

In temperate climates, stem growth occurs during the warm spring and summer months. The cells formed in spring when active growth is taking place are characteristically large, while during the summer only smaller cells are added. This alternation of cells results in the formation of definite concentric rings readily seen in cross sections of woody stems. Actually the growth increment is in the form of a sheath continuous over the entire stem except at the growing tips. External conditions may have a profound effect on the appearance of the annual ring; favorable growing seasons with ample moisture result in broad rings, while seasons of drought produce narrow rings. Removal of surrounding over-shading trees may result in a pronounced increase in the thickness of the annual ring. At times events such as severe defoliation by insects or cases of drought may produce two rings in one season; such rings are ordinarily not sharply distinct as are normal ones, and are called false annual rings. Counting of annual rings gives an accurate index of the age of the tree, while attention to details such as variable thickness of successive rings serves to indicate environmental changes. By careful comparison of different logs, even though they be largely reduced to charcoal, one may determine the actual year in which the ring was formed. By this means it has proved possible to establish the probable age of many ruins in the southwestern states. In tropical countries having a continuous growing season, annual rings are not formed or are only slightly developed. If, however, alternating rainy and dry seasons occur, then they appear.

ANNULUS. In the sporangium of many ferns, a ring of cells which have their walls characteristically thickened, and bring about the violent discharge of the spores within. In agarics, the ring of tissue which is found around the stalk in many genera, is also known as an annulus. See accompanying figure.



Fern sporangia: (1) unopened; (2) discharging spores; (3) empty.

ANNULUS (Geometry). See **Circle.**

ANODE. In the most general sense, an anode is the electrode via which current enters a device. The anode is the positively charged electrode of an electrolytic cell. See **Electrochemistry**. The anode (also frequently called the plate) is the principal electrode for collecting electrons in an electron tube, and is, therefore, operated at a positive potential with respect to the cathode.

ANODE SHEATH. In a gas discharge tube, the electron boundary which exists between the plasma and anode when the current demanded by the anode circuit is larger than the random electron current at the surface of the anode.

ANODIC OXIDATION. Since oxidation is defined not only as reaction with oxygen, but as any chemical reaction attended by removal of electrons, then when current is applied to a pair of electrodes so as to make them anode and cathode, the former can act as a continuous remover of electrons and hence bring about oxidation (while the latter will favor reduction since it supplies electrons). This anodic oxidation is utilized in industry for various purposes. One of the earliest to be discovered (H. Kolbe, 1849) was the production of hydrocarbons from aliphatic acids, or more commonly, from their alkali salts. Many other substances may be produced, on a laboratory scale or even, in some cases, on an economically sound production scale, by anodic oxidation. The process is also widely used to impart corrosion-resistant or decorative (colored) films to metal surfaces. For example, in the anodization or Eloxal process, the protection afforded by the oxide film ordinarily present on the surface of aluminum articles is considerably increased by building up this film by anodic oxidation. Also, one process for coloring the surface of aluminum, and retaining a metallic luster, is by adding substances to the metal, and subsequently oxidizing the surface anodically.

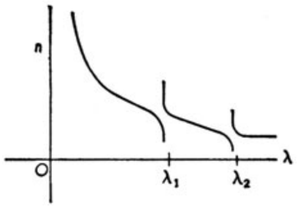
ANODIZE. This term means to place a protective film on a metal surface by electrolytic or chemical action in which the metal surface is made the anode in an electrochemical process. Aluminum and magnesium parts of electronics equipment are frequently anodized.

ANOLIS (*Reptilia, Sauria*). The name of a genus of lizards adopted also as a common name; small, mostly brightly colored lizards of the warmer latitudes of the Americas. The little lizard sometimes sold under the name chameleon is the Carolina anolis, *Anolis carolinensis*, and is a common species of the southern United States and southward.

ANOMALODESMACEA. An order of bivalve mollusks, mostly burrowing marine species.

ANOMALOUS DISPERSION. Ordinarily the refractive index n of a medium decreases with increasing wavelength λ (see **Dispersion**). It often happens, however, that in the immediate vicinity of a certain wavelength λ_1 there is a break or discontinuity in the dispersion curve and the usual rule may be locally reversed (see figure). In some cases there are several such points, $\lambda_1, \lambda_2, \lambda_3, \ldots$. These discontinuities correspond to lines or bands in the absorption spectrum of the medium. In the Sellmeier equation

$$n = 1 + \frac{A\lambda^2}{\lambda^2 - \lambda_1^2} + \frac{B\lambda^2}{\lambda^2 - \lambda_2^2} + \ldots$$



Variation of refractive index with wavelength, illustrating anomalous dispersion.

the several fractional terms make provision for the respective discontinuities. The absorption wavelengths $\lambda_1, \lambda_2, \ldots$ and the constants $A, B, \ldots$ must be determined experimentally. It will be noticed that n becomes infinite at every wavelength λ_k and is finite at all other wavelengths. If there is pronounced absorption and anomalous dispersion in the visible range, the medium appears colored, as illustrated by transparent dyes.

ANOMALY (Oceanography). The difference between the conditions actually observed at a particular point of measurement and an ocean of standard or arbitrary temperature and salinity.

ANOPLURA. The order of insects which includes the true or sucking lice. They are wingless parasitic insects with mouths formed for piercing and sucking. See **Louse.**

ANOREXIA. A loss of appetite or distaste for food, present for short periods in connection with a variety of diseases, is known as *anorexia.* Anorexia and weight loss frequently occur among patients with cancer and also arises in chronic alcoholism. The conditions may be due directly to the presence of the tumor in some cases. Anorexia may arise because of the production of anorexigenic peptides, a negative nitrogen balance as the result of unreutilized tumor amino acids, an uncoupling of oxidative phosphorylation, or a degraded glucose tolerance. Frequently, anorexia and weight loss are manifestations of the psychological and emotional stresses of a malignancy. Anorexia is almost universally present in subacute bacterial endocarditis. Usually where anorexia arises from a physical disorder that can be cured or arrested, the patient's appetite will ultimately return.

In some animals, particularly those species that hibernate, anorexia is naturally programmed. Fasting during hibernation has been recorded for centuries. However, more recent studies have shown that some animals eat very little and lose weight even when food is available. Often, this occurs when there are other more important activities which compete for the animal's time and attention. An example is that of bull seals that go without feeding for many weeks while minding their duties of defending territory and harem. Fasting associated with incubation, migration, and molting, as well as hibernation, also have been documented over a period of years. Some authorities believe that possibly through hormonal controls some species preprogram periods of anorexia that are made to coincide with other natural actions of greater importance at certain times for survival of individuals and preservation of species. In studies of this behavior, quite limited to date, no information has been uncovered that may be helpful toward understanding and treating the serious human disorder *anorexia nervosa*, which is described next. However, some early keys to improved control of obesity among humans are beginning to unfold.

Anorexia Nervosa. This disease has probably been best described as a complex psycho-endocrine disorder. This is a serious disease, with fatility rates ranging from 2 to 21%, as estimated by various authorities with death in many cases being suicidal. Currently the disorder is poorly understood and statistics are not fully reliable or representative. For example, the rate of incidence of the disease is not accurately known.

Anorexia nervosa occurs in young women and teenage girls at ten times the rate found in males of comparable age. In 1972, Feigner and associates developed the following criteria for diagnosis of anorexia nervosa: (1) loss of 25% of body weight, (2) a desire to lose weight, (3) an onset prior to age 25, (4) presence of amenorrhea (absence of menstruation), (5) overactivity, and (6) absence of known medical or psychiatric illness. Universally present is a fear of gaining weight. The fear leads to unusual eating patterns and the avoidance of foods of high caloric value. Amenorrhea may be present at the outset or develop during the course of the disorder. Although the illness includes an endocrine disturbance, it is specific and confined to the hypothalamic-anterior pituitary-gonad axis. In the absence of hypothalamic defects, the general view is that endocrine changes follow rather than lead the psychologic and emotional oddities of the disease. The disease has not been established as having an endocrine origin, but does cause serious endocrine consequences.

Fatness is intolerable to such patients. Sometimes a precipitating experience, difficult to comprehend by others, may be attributed by the patient as the trigger of the motivation to be thin. The loss of weight induced by the prolonged insufficient intake of carbohydrates is self-induced and thought to be a protective device. Dread of growing up, leaving home, becoming independent with aggravation by a domineering mother have all been associated with development of the disease which, although most common in the upper social classes, does not exempt any social strata. Most patients with anorexia nervosa are not considered psychotic. Sometimes, however, anorexia nervosa is part of the complex in a true mental illness. Patients tend to be normal in their approach to life excepting the topics of food and weight. It is not unusual for such patients to prepare nutritious meals for the consumption of others. Although the distinctive features of anorexia nervosa are easy to detect, examinations generally reveal few abnormalities except those of an endocrine nature. There are atypical cases which do make diagnosis difficult. Some of the aforementioned criteria may be absent or only partly present.

Therapy includes brief psychotherapy, isolation from family, hospitalization with tube feeding, and psychoanalysis, the course of treatment taking some two to three years with many opportunities for relapse. Numerous drugs have been used, but reliable comparative studies on their effectiveness remain to be developed. Hormonal substitution therapy has been used with limited success.

Recovery is uncertain and varies with individuals. Limited statistics indicate that about 48% of patients ultimately recover, including a return of menses, normal weight, and improved mental and psychosexual outlook. Intermediate results have been reported in 30% of cases; very poor results in 20% of cases.

In patients with hypothalamic defects, the most frequent mechanism of secondary amenorrhea is the absence of the LH (leutenizing hormone) required for ovulation. One of the most vulnerable components of the female reproductive endocrine system, the LH surge is inhibited in anorexia nervosa.

Delayed puberty may result from prepubertal anorexia nervosa. Persons with anorexia nervosa, as well as persons who have voluntarily (martyr syndrome) or involuntarily starved may develop leukopenia and neutropenia—low blood counts of white cells and neutrophils, respectively.

Additional Reading

Adams, R. D., and M. Victor, Editors: "Principles of Neurology," McGraw-Hill, New York, 1989.

Anderson, A. E.: "Males with Eating Disorders," Brunner/Mazel, New York, 1990.

Grunfield, C., and K. R. Feingold: "Metabolic Disturbances and Wasting in the Acquired Immunodeficiency Syndrome," *N. Eng. J. Med.*, 329 (July 30, 1992).

Rapaport, J. L.: "The Biology of Obsessions and Compulsions," *Sci. Amer.*, 83 (March 1989).

Rowland, L. P., Editor: "Merritt's Textbook of Neurology," 8th Edition, Lea and Febiger, Philadelphia, Pennsylvania, 1989.

Rusting, R.: "Starvaholics? (Anorexics May be Addicted to a Starvation 'High')," *Sci. Amer.*, 36 (November 1988).

Waldinger, R. J.: "Psychiatry for Medical Students," 2nd Edition, American Psychiatric Press, Washington, D. C., 1990.

Yates, A.: "Compulsive Exercise and the Eating Disorders: Toward An Integrated Theory of Activity," Brunner/Mazel, New York, 1991.

ANORTHOSITE. The name anorthosite was given by T. Sterry Hunt to rocks of gabbroid nature which were essentially free from pyroxene, hence almost wholly plagioclase *usually* labradorite. The term is derived from the French word for plagioclase, anorthose. Small quantities of pyroxene may be present as well as magnetite or ilmenite. The rock is commonly white to gray, bluish, greenish, or perhaps nearly black. A variety from the Province of Quebec is purplish-brown due to the inclusion of ilmenite dust within the feldspars. Although not a common rock in the ordinary sense of the word, occurrences of great areal extent are known in Canada, Norway, and Russia and in the United States in northern New York State and Minnesota. Opinions as to the origin of this rock differ. The development of anorthosite may have been due to the settling out of labradorite crystals from a gabbro magma as many believe, or there may have been an original anorthosite magma.

A study of anorthosite occurrences brings out two very curious circumstances, first, that there is no extrusive (lava) equivalent of anorthosite, and second, that most anorthosite masses seem to be of pre-Cambrian age.

ANOXEMIA. Deficiency in the oxygen content of the blood. This may be reduced:

1. When there is insufficient oxygen available to saturate the hemoglobin; this may occur in pulmonary diseases in which inflammatory processes interfere with the passage of oxygen into the blood, or under conditions such as are encountered in high altitude climbing, inhalation of inert gases (methane, helium), in which there is insufficient oxygen in the inhaled air, and in the late stages of cardiac and respiratory disease when insufficient air is inspired (anoxic anoxemia);
2. When the amount of hemoglobin in the blood is insufficient to carry the amount of oxygen required, as occurs in anemia from any cause, especially that following acute or chronic blood loss (anemic anoxemia).

Rapidly developing anoxemia leads, when the oxygen tension of the inspired air (normally 21%) falls to 10%, to cyanosis and to increased rate and depth of breathing; at 5% to loss of consciousness and ultimate death.

Slowly developing anoxemia may be compensated for by a process of acclimatization, not however without damage to vital tissues when the oxygen deprivation is prolonged.

ANSERIFORMES (*Aves*). A large number of goose-like or fowl-like birds which live near water and at least temporarily go into water are grouped in this order. They are ground and waterbirds; the length is 28–170 centimeters (11–67 inches), and the weight is 200–13,500 grams (7 ounces to 30 pounds). The nostrils connect with one another, and the lower mandible has a long process at the angle. The sternum has two indentations or two foramina at the rear; these are absent in the fossil giant duck (*Cnemiornis*) of the glacial period. Two pairs of muscles are located between the sternum and the trachea. The neck is extended in flight. There are 10–11 primaries, the fifth secondary is absent (diastataxic wing), and there are 12–24 tail feathers. Many down feathers are found in the fully developed plumage. The unspotted eggs are light in color. The young are nidifugous, have a dense downy plumage, and are tended for a long time by one or both parents. They are distributed over all continents except Antarctica. The *Anseriformes* are divided into two families distinguished by the absence or presence of horny lamellae in the beak: (1) screamers (*Anhimidae*), and (2) ducks and geese (*Anatidae*).

The screamers (family *Anhimidae*) are almost goose-sized birds of fowl-like appearance, with fairly thick, long legs and feet without webs. The weight is 2–3 kilograms (4.4–6.6 pounds).

There are two genera with marked differences in their internal structure: the horned screamers (*Anhima*), with 14 tail feathers, and the crested screamers (*Chauna*) with 12 tail feathers. Altogether there are three species with no subspecies: (1) the horned screamer (*Anhima cornuta*) which reaches a length of 80 centimeters (31 inches) and inhabits the flood forests of the Amazon delta, (2) the crested screamers (*Chauna torquata*) with a length of 90 centimeters (35 inches), and are found in swampy pampas areas of the La Plata States, and (3) the black-necked screamer (*Chauna chavaria*), with a length of 70 centimeters ($27\frac{1}{2}$ inches), is found on forest rivers of Colombia and Venezuela.

Outside the breeding season, the horned screamers live in troops of 5–10 birds. The crested screamers are, however, found in larger flocks which circle above the waters in their habitats in the evenings, calling melodiously. In contrast to the *Anatidae*, they can glide well. In spite of their unwebbed feet, screamers swim very well; the crested screamer will even climb onto the leaves of floating plants from the shore. These birds calmly walk about the shore or in shallow water; their food is entirely vegetarian and they obtain some of their food while swimming. They readily perch on the branches of trees and, when disturbed or pursued, generally take to trees.

All other members of this order are included in the family of ducks and geese (*Anatidae*). They have horny lamellae on the interior of the beak near the cutting edge. There are webs between the anterior toes. The upper mandible has a particularly hard process, the "nail," at its tip. There are large nasal cavities; because of this, geese breathe faster when reacting to olfactory stimuli. There are 16–25 cervical vertebrae. They cannot soar or glide to any extent; a few species are flightless, while others have a rapid flight.

There are three subfamilies: (1) the magpie goose (*Anseranatinae*) with only minute webs, (2) the geese and relatives (*Anserinae*) with larger webs and small scales on the tarsus and toes, and (3) the ducks and their relatives (*Anatinae*), which also have large webs.

The *Anatidae* are made up of many species that are colorful and of many shapes. There is a wide range of changing forms and ways of life, from the minute African pigmy geese to the trumpeter swan, which weighs $13\frac{1}{2}$ kilograms (30 pounds), and from the inconspicuous greylac goose to the colorful plumage of the king eider. Nevertheless, the different species have common characteristics which justify their grouping in one family. Thus all are water birds, even species which live mainly on land like the Cape Barren goose and the Hawaiian goose. Since, as swimmers, their plumage must always be greased, they have a particularly large preen gland.

All *Anatidae* have webbed feet. The webs extend from the second to the third and from the third to the fourth toe. *Anatidae* do not have to make any particular effort to remain on the surface of the water or to swim. Their buoyancy is due mainly to the air held in the plumage. They take great care not to get water under their plumage. The closed wing is covered by feathers projecting up from the side of the breast so that, of the wing feathers, only the scapulare and the primaries are generally exposed. The body plumage is continually and carefully covered with oil from the preen gland and so forms a layer impermeable to water.

A particular adaptation to the requirements of swimming is the broad cross section of the body of most *Anatidae*. This broad, bargelike body maintains its balance despite wind or waves. Another adaptation is the shortening of the thigh and tarsus. The tarsus functions like the arm of a lever in swimming, and in slow swimming is almost the only part of the leg that moves. Only in faster swimming is the thigh involved as well; it is drawn back with partially extended knee, the lower leg serving merely to transmit power.

To reduce resistance to the water when the tarsus moves forward, the webs and toes are folded together and the toes are bent. In pushing back, the toes and webs are fully extended and form an effective oarlike surface. As in walking, so in swimming the legs move alternately. Only the mute swan in its aggressive display swims with jerky movements, pushing back with both legs at the same time; however, it swims no faster by this method.

According to their manner of obtaining food, the *Anatidae* can be divided into several groups. Swans, shelducks, and surface feeding ducks "up end"; that is they immerse the head and neck with the rear of the body projecting almost upright above the water. In this way, they can feel over the bottom of shallow waters with the beak, and obtain food by straining it out of the water. Diving ducks also generally get their food from the bottom, but they reach greater depths and dive completely below the surface. Lastly the mergansers chase fish beneath the surface.

The true geese, swans, and whistling ducks are entirely vegetarian. Mergansers, scoters, eider ducks, and the South American torrent ducks take only animal food. The rest take both plant and animal food. The amounts of food taken are at times considerable. In the digestive tract of an eider duck, 114 mussels were found, some of which were already partially digested within their shells; the gullet and stomach of a velvet scoter contained 45 oysters.

Most of the species build their nests on the ground. Some of the genus *Tadorna*, like the common shelduck, prefer burrows in the ground as nest site, although some other species of ducks, among them the mallard, also nest on trees. The Orinoco goose, many of the *Cairini* like the maned wood duck, the mandarin, and the wood duck, the Brazilian teal, the pigmy geese, the comb duck and its relatives, the goldeneyes, and some mergansers breed preferably in tree cavities. Nests of the magpie goose, the coscoroba swan, and many diving ducks and stiff-tailed ducks are often found in dense swamp vegetation on the water.

The nest construction is simple. As far as it is possible, the birds make a hollow in the ground and pull in stems and leaves from around the nest site, as far as they can reach with their outstretched necks. All species cover incomplete clutches with plant material when leaving the nest. Shortly before the last egg is laid, the females pluck the nest-down and line the nest with it. See also **Poultry; Screamer;** and **Waterfowl.**

ANT (*Insecta, Hymenoptera*). Social insects of varied structure and habits. They may be distinguished from the related bees and wasps by the form of the slender petiole which connects thorax and abdomen; in the ants it is expanded above and looks wedgelike in profile. Ants have been known to exist on earth for some 30 to 40 million years. Evidence of their early existence is found in historic fossil Baltic amber. The average ant is about one-sixteenth inch in length, but the large Texas ant may measure up to one inch or more. Stages in the development of the ant are egg, larva, pupa, and adult. Most ants are omnivorous and wingless. See Fig. 1. There are over 5000 species of ants. Among varieties found in the United States are the army ants, carpenter ants, the leaf cutters, the dairyman, the garden (common ant), honeydew, and mound-building ants.

Fig. 1. Generalized silhouette of an ant.

A sampling of ant types and habits would include: the voracious Argentine ant which steals eggs and ants from other nests, making the captive ants their slaves; the honey-pot ant which stores extra honey in its flexible body to furnish food for the young; the Legionaries ants which travel in single file to steal and store other insects for food, which live in wet areas mostly under foliage and are nearly blind; the garden ant which builds small mounds of sand in gardens or on walkways, brown in color, and with a good-natured temperament; the Texas ant known as the fungus grower in the ant world, and which builds mounds up to 12 inches (0.3 meter) in diameter, and chews leaves into a mulch for the growth of fungus which then becomes its food; the harvest ant or seed collector, one of the largest of the ants, which stores small seeds of various kinds in its mound and whose bite is painful; the carpenter ant which drills tunnels in rotting wood; the sugar ant which is extremely small, very light-brown in coloration, harmless, and which is excessively fond of sweets; the tree ant of India (*Olcophylla smaragdina*) which uses its larvae as a means of sewing leaves together to make a nest; Brazil's terrible ant which is fully 1 inch (1.5 centimeters) long, and which produces a vicious sting.

It has been determined that the behavior habits of ants are strictly acquired naturally, as in the case of the honeybee, and that no intelligence in the normal interpretation of the word is required. Ants recognize excitement and unusual conditions, however, and relay this information to other ants by stroking them with their antennae, pecking them on the head or thorax and, if a danger signal is required, the ant will open its jaws wide, or it may hold its abdomen high and run about wildly. Ants are also known to leave scent trails by pressing their stomach close to the ground. Although the scent persists for just a short time, the scent trail is refreshed by other ants repeating the same procedure.

Texas Leaf-Cutting Ant (*Atta Texana,* Buckley). Native to southern and eastern Texas (and western Louisiana), this species of ant prefers loamy soil that is adequately drained. Some nests range from 10 to 20 feet (3 to 6 meters) in depth, incorporating numerous craters, but not rising very high above ground level. Such nests are the habitat of many thousands of ants. The worker ants of this species range from $\frac{1}{16}$ to $\frac{1}{2}$ inch (1.5 to 12 millimeters) in length, they are light-brown in color, and have hardened bodies equipped with many spines on the head and thorax. Food for these ants is prepared by the workers who macerate freshly cut leaves which subsequently promotes the growth of a fungus used as food. In addition to serious pests in the garden, the ants attack field crops and, in particular, they are damaging to young pine seedlings and thus must be eradicated in areas of reforestation. Poisoned bait which the ants can carry back to their nest is effective. Argentine bait has been successful. This includes tartaric acid crystals, benzoate of soda, and sodium arsenite. These materials are boiled, along with sugar, and then mixed with honey. The resulting sirup is used as bait for ants that like sweet foods, such as the Argentine ant. The bait does not work, however, with the Pharaoh ant, the southern fire ant, or the tiny thief ant. A similar bait for protein-loving ants is prepared from thallous sulfate, groundnut (peanut), butter, and German sweet chocolate. Both of these concotions are quite poisonous and must be prepared and used with extreme caution, marking all containers as *poison*. Placing insecticides in the normal pathways of travel of ants also is effective. Pouring liquid carbon bisulfide directly into nest openings also can be effective.

Imported Fire Ant (*Solenopsis saevissima richteri,* Forel). Native to South America, the fire ant invaded the United States at Mobile, Alabama in 1918. Since then, it has spread into more than 130 million acres (52 million hectares) in Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, and Texas. This is a small, aggressive insect that produces a painful, burning sting. When disturbed, the ant is quick to attack both people and domestic animals. Each colony of imported fire ants builds a hard-crusted nest, or mound, sometimes 3 feet (1 meter) high and nearly 3 feet (1 meter) across. In some areas, there may be as many as 50 mounds per acre (125 per hectare), making it difficult to operate mowers and other machinery in pastures and fields, as well as lawns and park grounds. See Fig. 2.

Fig. 2. Areas of infestation of the imported fire ant in the southeastern United States as of 1990. Migration of the ants commenced in a comparatively small strip of land along the southern borders of Louisiana and Mississippi. The infested area spread southwesterly to include part of Texas and northeasterly to include nearly all of Alabama, Georgia, Florida, and South Carolina. Entomologists forecast the spread ultimately will include nearly all of central and south Texas, southern Arizona, and north along the western half of California, Oregon, and Washington. A meticulous survey of over 29,000 physicians in the affected areas (conducted in 1989) identified 32 deaths attributed to anaphylaxis caused by fire-ant stings. The patients who died ranged from 16 months to 65 years of age and usually had been stung fewer than five times. Anaphylaxis may occur hours after a sting. Seizures and mononeuritis also have been reported after fire-ant stings. (Map: *USDA*.)

The damage from the pest is difficult to measure in economic terms. The stings cause blisters that require as long as 10 days to heal. If the blisters break, infection may develop. Some people have been hospitalized; a few have died, primarily from allergic reaction to the stings. Some farm workers may refuse to work on land where these ants are numerous for fear of being stung while clearing clogged mower blades, handling crops, and performing other agricultural tasks.

Imported fire ants look like ordinary house and garden ants. They are from $\frac{1}{8}$- to $\frac{1}{4}$-inch (3 to 6 millimeters) long and reddish-brown or dark brown to black in color. A single mature mound contains a queen ant, several thousand winged males and females (future queens), and up to 100,000 workers. The winged forms leave the mound, most frequently in May and June, and mate in flight. Afterward, the queens land and break off their wings. They dig shallow burrows in the soil and begin to lay eggs that start new colonies. Winds and air currents may carry the new queens 12 or more miles (22+ kilometers) during mating flights. They prefer to build their "homes" in open, sunny areas. Thus, the most valuable land on farms and in suburbs is often the most likely to be infested.

Cooperative programs between federal, state, county, and local governments have been established to make continuous surveys and to provide controls over the pest. Quarantines may be established to prevent further widespreading of the insect.

Medical Implications of Fire-Ant Stings. The human resistance to the effects of fire-ant stings ranges widely. Some persons develop extreme sensitivity (anaphylaxis) to a single sting. In relatively infrequent cases, seizures and mononeuritis have been reported after fire-ant stings.

Investigators have found that the venom of the fire ant is quite different from that of wasps, bees, and hornets, which are aqueous solutions containing proteins. Fire-ant venom, in contrast, contains up to 95% alkaloids, with only a small aqueous fraction of soluble proteins. Researchers have found that almost 100% of the fire-ant venom contains 2,6-di-substituted piperidines that have hemolytic, antibacterial, insecticidal, and cytotoxic properties. However, the very small quantity of proteins present in the venom induce allergic responses in some individuals.

As reported by R. D. DeShazo (University of South Alabama College of Medicine), B. T. Butcher (Arthritis Foundation, Atlanta, Georgia), and W. A. Banks (USDA, Gainesville, Florida), "Since the natural history of sensitivity to imported fire ant stings is unknown, the indications for immunotherapy to prevent the recurrence of sting-related anaphylaxis are unclear. Studies of anaphylaxis in response the stings of other insects, such as honeybees and yellow jackets, have established that the production of IgE (antibody responses) after stings is common and often transient. Moreover, there are differences between adults and children in the clinical manifestation of anaphylaxis, the probability of serious reactions after subsequent stings, and the duration of sensitization. This information is useful in selecting patients who would benefit from immunotherapy with honeybee and yellow-jacket venom. Unfortunately, similar information is not available for imported fire-ant venom."

Fire-ant stings produce at least three types of local reactions: (1) a wheal-and-flare reaction; (2) a sterile pustule; and (3) a large local reaction. The wheal-and-flare reaction usually resolves in about a half-hour to an hour and evolves into a fully developed sterile pustule at the site of the sting within 24 hours. The epidermis covering the pustule sloughs off over a period of 48 to 72 hours. Healing occurs at the base of the lesion, where it is covered with new epidermis. No treatment has been found to prevent or resolve pustules. If there are several pustules and they become excoriated (i.e., if they are worn off the skin) and then superinfected, pyoderma and even sepsis can result, a condition of specific concern to diabetics. Much more detail is provided in the DeShazo, Butcher, and Banks reference listed.

Early, rather effective control chemicals such as the pesticide heptachlor and later mirex were used in attempts to eradicate the fire-ant mounds. These substances were banned because of their carcinogenicity. Drenching individual ant mounds with diazinon or injecting mounds with chlorpyrifos (under pressure) are less effective, because adjacent areas soon become infested from the perimeter. In recent years, there have been numerous reports of how fire ants are attracted to an electrical field. They have eaten through electrical lines, causing power outages. Incidentally, the Texas site for the Super Conducting Supercollider was found to be infested with fire ants. The cause of the electrical field attraction of the insects has not been explained.

Cornfield Ant (*Lasius alienus,* Förster). The cooperation between ant and aphid is mentioned in the entry on **Aphid.** The corn root aphid (*Anuraphis maidiradicis*, Forbes) and the cornfield ant provide an excellent example of this cooperation and thus, indirectly, the cornfield ant contributes in a major way to the damage wrought on corn by the corn root aphid. Whenever the food producer notices half-grown corn (maize) that is not doing very well, as indicated by sluggish growth and yellow-or red-tinged leaves, it will be noted that, at the base of the plant, there will be numerous blue-green aphids and a great activity of small brown cornfield ant milling about between the corn roots (which they tunnel) and a number of anthills near the plant. During the winter season, aphid eggs are collected by the ants and stored in their nests during the cold weather. The eggs are frequently moved about to assure they are obtaining the best exposure to temperature and humidity. In early spring, the aphid eggs commence to hatch. At this time, the ants carefully carry the tiny eggs to the roots of nearby smartweed plants or other grasses suitable for aphid feeding. After 2 to 3 weeks of this feeding process, the full mature aphid females commence to give birth to live female young. In turn, these females mature and give birth to others in 2 weeks or less. Reproduction is by parthenogenesis (development of the egg without fertilization). Throughout the summer, the ants care for the welfare of the aphids. Distribution of the aphids in a corn (maize) field, for example, depends almost totally upon the actions of the ants. The reward to the ant is the honeydew, a sticky sweet exudation from the anal opening of the aphid. The honeydew comprises a major part of the ant's diet. It has been observed that ants have moved over 150 feet (50 meters) from a grass meadow to a cornfield, moving their own young as well as large numbers of aphids, which the ants then turn out to pasture on young corn roots.

Thief Ant (*Solenopsis molesta,* Say). A common household pest, the thief ant prefers protein foods to sweet foods, and sometimes damages grain seed when germinating.

Additional Reading

Burnham, L.: "High Rate of Return (Fire-Ant Sperm Banks Yield One Worker for Every Three Sperm)," *Sci. Amer.*, 22 (December 1988).

Byrne, G.: "The Fire Ants," *Science*, 32 (January 6, 1989).

DeShazo, R. D., Butcher, B. T., and W. A. Banks: "Reactions to the Stings of the Imported Fire Ant," *N. Eng. J. Med.*, 462 (August 16, 1990).

Hall, A. J.: "Scourge of the Forest, Moundbuilder Ants," *Nat'l Geographic*, 812 (June 1984).

Handel, S. N., and A. J. Beattle: "Seed Dispersal by Ants," *Sci. Amer.*, 76 (August 1990).

Haskins, C. P.: "The Ant and Her World," *Nat'l. Geographic*, 774 (June 1984).

Holldobler, B.: "Ways of the Ant," *Nat'l. Geographic*, 778 (June 1984).

Holldobler, B.: "Communication Between Ants and Their Guests," in *Life at the Edge*, 122, Freeman, New York, 1988.

Holldobler, B., and E. Wilson: "The Ants," Harvard University Press, Cambridge, Massachusetts, 1990.

Moffett, M. W.: "Marauders of the Jungle Floor," *Nat'l Geographic*, 272 (August 1986).

Moffett, M. W.: "Trap-Jaw Ants," *Nat'l. Geographic*, 394 (March 1989).

Weiss, R.: "Ants Get a Transforming Charge," *Science News*, 412 (December 23, 1989).

ANTACIDS. These are formulations widely used in the treatment of excessive gastric secretions and peptic ulcer. Several factors determine the efficacy of antacids, including (1) the ability and capacity of the stomach to secrete acid; (2) the duration of time the antacid is retained in the stomach; and (3) the nature of the gastric response upon eating.

Five principal active ingredients are used in antacid preparations: (1) *Sodium bicarbonate* is a rapid and effective neutralizer. The compound does yield large amounts of absorbable sodium, undesirable in some persons (heart disease; hypertension). The compound also may induce milk-alkali syndrome. (2) *Calcium carbonate* is a strong, effective neutralizer, but can cause constipation, hypercalcemia, acid rebound, and milk-alkali syndrome. (3) *Aluminum hydroxide* provides slow and not potent action. The compound causes constipation, absorbs phosphates, as well as certain drugs, such as tetracyclines. (4) *Magnesium hydroxide* which provides a slow and prolonged action with no major side reactions. (5) *Magnesium trisilicate* which acts like magnesium hydroxide, but which is poorly absorbed and acts as an osmotic laxative. In cases of renal insufficiency, the serum magnesium should be monitored.

The foregoing compounds are frequently used in combination and, in some, simethicone is added to relieve flatulence. There are striking differences of commercial antacids in terms of their neutralizing capacity.

The physician is concerned with at least three factors when prescribing antacids: (1) Acid rebound (associated with calcium carbonate); (2) milk-alkali syndrome (caused by ingestion of large quantities of alkali); and (3) phosphorus depletion (by aluminum salts). The mechanism of acid rebound, especially in the long-term use of calcium carbonate, is poorly understood. It has been established that there is an excessive reacidification of the antrum (pyloric gland area) a number of hours after ingestion of calcium carbonate.

The ingestion of a quart of milk or more while taking large amounts of alkali, as from antacids, sets up conditions favorable to milk-alkali syndrome. Generally, with withdrawal of the milk or the antacid, the condition is self-correcting. Symptoms of milk-alkali syndrome include nausea, vomiting, anorexia, weakness, polydipsia, and polyuria. Abnormal calcifications also may occur in the chronic stage and other symptoms include mental changes, asthenia, aching muscles, band keratopathy, and nephocalcinosis. Symptoms of milk-alkali syndrome sometimes tend to mimic hyperparathyroidism and vitamin D intoxication.

ANTARCTIC CONVERGENCE. A distinct, natural oceanographic boundary around the continent of Antarctica. The boundary is more or less equivalent to the 50°F (10°C) isotherm for the warmest month. It has been determined that the colder, denser Antarctic waters sink sharply below the warmer, lighter Subantarctic waters with very little mixing. The flora and fauna reflect the water and air temperature differences on either side of the boundary.

ANTARCTIC REGION RESEARCH. See **Polar Research.**

ANTARCTIC WATERS. Water masses in or associated with the Antarctic Ocean, including:

Antarctic Bottom Water. An oceanic water mass arising close to the margins of the Antarctic continent. It is particularly dense due to the cold and to the fact that the surface waters freeze over in winter, leaving the salt behind to increase the density of the remaining water. Sinking to the bottom of the ocean, the water creeps north until it encounters the North Atlantic Deep Water, with which it is believed to merge.

Antarctic Circumpolar Water. Also called West Wind Drift, this is the oceanic water mass with the largest volume transport (approximately 110×10^6 cubic meters per second) ($3{,}883 \times 10^6$ cubic feet per second) and the swiftest current. It flows from west to east through all the oceans around the Antarctic continent. The flow is locally deflected from its course, particularly by the distribution of land and sea and partly by the submarine topography. Besides the bends that are associated with the bottom topography, the effects of the distribution of land and sea and of the currents in the adjacent oceans are also evident. On its northern edge, it is continuous with the South Atlantic current, the South Pacific current and the eastward-flowing extension of the Agulhas current in the Indian Ocean. Salinity maxima occur at depths ranging from 700–1300 meters (2,310–4,290 feet), averaging 34.8%. The temperature range is 0 to 2°C (32° to 35.6°F). It is surface water in some regions; a deep mass in others.

Antarctic Intermediate Water. This is the oceanic water mass originating in the northern part of the Antarctic closest to the equator. The water moves northward as a surface current until it meets the warmer waters of the South Atlantic Ocean. Because it is colder and thus heavier, the Antarctic water sinks below the surrounding warm water to a depth of approximately 600–900 meters (1,980–2,970 feet) and continues northward until it returns to the surface between 20° and 30° north latitude.

Antarctic Surface Water. A relatively shallow oceanic water mass extending from the Antarctic convergence, where it meets the Sub-antarctic Water, to the shores of Antarctica. Its depth increases from about 80 meters (264 feet) in the Atlantic and 150 meters (495 feet) in the Pacific to 300–400 meters (990–1,320 feet) as it nears Antarctica. Temperature range is from −2 to 3.5°C (28.4° to 38.3°F); salinity from 32.8% to 34.5%.

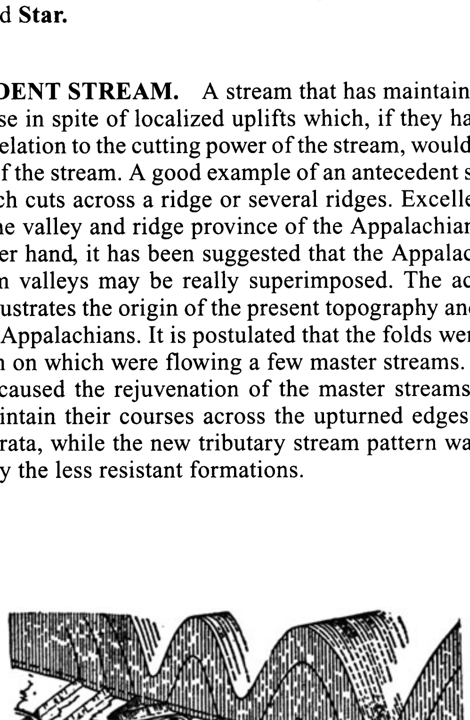
Subantarctic Water. An oceanic water mass extending on the surface from the South Atlantic Central Water and the South Pacific Central Water to the Antarctic Surface Water in the south. In the Atlantic, it extends only from about 52° to 53° south latitude (from the subtropical convergence to the Antarctic convergence). In the Pacific Ocean, it covers a much larger area. Temperature range at the surface is 3 to 10°C (37.4° to 50°F); salinity from 33.8% to 34.8%.

ANTARES (α *Scorpii*). A star, whose name is derived from two Greek words signifying that it is "similar to" or the "rival of" Mars, doubtless because of its distinctly reddish hue. In fact, this reddish color has always made Antares an object of interest and importance in the ancient religions, and many of the Egyptian temples are so oriented as to indicate that this star played an important part in their ceremonials. Antares was one of the four royal stars of the Persians about 3000 B.C., and some writers claim that it is the "lance star" referred to in the 38th chapter of the Book of Job.

The diameter of Antares has been determined with the stellar interferometer and found to be about 7.5×10^8 kilometers, or slightly greater than the distance of Mars from the sun. It is a typical M spectral-type giant star of very low density.

Ranking sixteenth in apparent brightness among the stars, Antares has a true brightness value of 5,000 as compared with unity for the sun. Estimated distance from the earth is 400 light years. See also **Constellations;** and **Star.**

ANTECEDENT STREAM. A stream that has maintained its consequent course in spite of localized uplifts which, if they had proceeded rapidly in relation to the cutting power of the stream, would have caused diversion of the stream. A good example of an antecedent stream valley is one which cuts across a ridge or several ridges. Excellent examples occur, in the valley and ridge province of the Appalachian Mountains. On the other hand, it has been suggested that the Appalachian antecedent stream valleys may be really superimposed. The accompanying diagram illustrates the origin of the present topography and stream pattern of the Appalachians. It is postulated that the folds were reduced to a peneplain on which were flowing a few master streams. Uplift of the peneplain caused the rejuvenation of the master streams which were able to maintain their courses across the upturned edges of the more resistant strata, while the new tributary stream pattern was largely determined by the less resistant formations.



Structural and erosional history of the Appalachian Range. (*W. M. Davis*)

ANTELOPE (*Mammalia, Artiodactyla*). The antelopes (*Antelopines*) comprise one of the larger groups of the order *Artiodactyla* (even-toed hoofed mammals). In anatomy, physiology, appearance, and habits, the antelopes lie between the oxen and the sheep and goats. Many species occur in Africa and some in India and Tibet. The Pronghorn Antelope of western North America belongs to a separate group (*Antilocaprines*). The horns of the pronghorn are hollow; the horns of true antelopes are almost solid. See **Pronghorn Antelope.**

The horns, in fact, are one of the most interesting features of the antelopes. The horns consist of a structure of bone covered with keratin which is harder than bone and grows out from the animal's skull. Horns of older rams grow into nearly complete circles. It has been reported that the noise from clashing horns during fights between the males of certain species of antelopes can be heard sometimes for a distance of two miles. The shape and size of the horns frequently serve as an excellent index of the subfamily or species. For example, the Kudii has large corkscrew horns; the Giant Sable, crescent-shaped horns; the Oryx, rapierlike horns, etc. Rings on the horns form yearly, but the age is often difficult to determine with accuracy by counting the rings because there is some overlapping.

As a general observation, the various species of antelopes can run fast, but not always fast enough to avoid death on the open plains where large carnivorous animals consider the antelope good eating. Most species of antelopes are quite alert looking, with head held high, ears erect, large eyes, short hair, graceful build, and frequently attractive and distinguished horns. Size ranges from about that of a goat to as large as a horse. Some antelopes have markings on the face and head which also are indicative of species. In several species, the coats blend in well with the surrounding habitat. Some people in antelope-inhabited areas consider antelope flesh as good meat, even a delicacy.

The general organization of the *Antelopines* group is given in the accompanying table. Only the most important species are indicated. The following descriptive paragraphs follow the general organization of the table.

With the exception of a few species, the antelopes are almost exclusive to Africa where they prefer savanna, grassland, scrub, and semidesert as a habitat. The term *antelope* is sometimes used loosely for the related bovines and caprines.

Oryx or besia antelope. (*American Museum of Natural History.*)

The Horse-Antelopes are large-hoofed animals with horse-shaped bodies. The Giant Sable antelope is known only in Angola and considered quite rare. Coloration is purplish-black. The rapier-horned antelopes are small and very horselike in appearance. See accompanying figure. The Oryx is found in the desert regions of Africa and thence to Syria. All species have very long horns, straight or slightly recurved.

GENERAL ORGANIZATION OF THE ANTELOPES
ANTELOPINES

HORSE-ANTELOPES (*Hippotraginae*)
Sabre-horned Antelopes (*Hippotragus*)
—Giant Sable Antelope (*H. tariani*)
—Common Sable Antelope
—Roan Antelope
Rapier-horned Antelopes (*Aegoryx* and *Oryx*)
—White Oryx (*A. algazel*)
—True Oryxes
—Gemsbock
—Beisa
—Beatrix Oryx
Screw-horned Antelopes (*Addax*)

DEER-ANTELOPES (*Alcelaphinae*)
Hartebeests (*Alcelaphus*)
—Coke's Hartebeest
Damalisks (*Beatrugus* and *Damaliscus*)
—Hunter's Hartebeest (*B. hunteri*)
—Korrigum
—Topi
—Sassaby
—Blexbok
—Bontebok
Gnus (*Connochaetes* and *Gorgon*)
—White-tailed Gnus
—Brindled Gnus
—White-bearded Gnus

MARSH-ANTELOPES (*Reduncinae*)
Waterbucks (*Kobus*)
Lechwes (*Onotragus*)
Kobs (*Adenota*)
—Kob
—Puku
Reedbucks (*Redunca*)
The Rhebok (*Pelea*)

BLACKBUCK (*Antilopinae*)

PIGMY ANTELOPES (*Neotraginae*)
Klipspringers (*Oreotragus*)
Oribis (*Ourebia* and *Raphicerus*)
—Steinboks
—Grysboks
Sunis (*Neostragus*)
The Beira (*Dorcatragus*)
Dik-Diks (*Madoqua Rhynchotragus*)
Royal Antelopes (*Nesotragus*)

GAZELLES (*Gazellinae*)
Impalla (*Aepyceros*)
The Gerenuk (*Litocranius*)
The Dibatag (*Ammodorcas*)
The Springbuck (*Antidorcas*)
The Addra (*Addra*)
True Gazelles (*Gazella*)
Goat-Gazelles (*Procapra*)

The screw-horned antelopes are natives of the Sahara. The species is known for going long periods with absolutely no water.

Several species of the hartebeest make up a significant portion of the group of Deer-Antelopes. Generally they have large, ringed horns, irregularly spiraled with the tips pointing back. These animals are graceful and the horns of some species are described as resembling a lyre. The animals are about 4 feet high (1.2 meters) at the withers (ridge between shoulder bones). The forequarters are heavier and much higher than the hindquarters. Possibly the first hartebeest to be recognized was the Titel or Bubal, which roamed across north Africa in the days of the Roman Empire and was frequently called a horned horse. This animal, smaller than most hartebeests and with relatively short thick horns, ringed and black in color, became extinct in the early 1900s. Another member of the hartebeest group is the Konzi, with small horns, pale color, and black tail and front of legs black. It is a very lively animal, living in small parties, often in company with zebra and waterbuck, usually in the flat wooded districts of Zambia and Rhodesia along the Zambezi. Another hartebeest is the Korigum of central Africa, also sometimes called the Senegal antelope. The Sassaby is sometimes called the bastard hartebeest. The animal stands about 4 feet high (1.2 meters), with horns up to 15 inches (38 centimeters) long. Coloration is deep red, blending into black on the back. Hunter's hartebeest is very rare, found only in the environs of southern Somaliland. The term *hartebeests* is often used in describing both hartebeests and the damalisks.

Gnus are sometimes called *wildebeests*. The animals generally have a large head, strong curved horns, an erect bristly mane and a bristly muzzle. The withers are high and the tail is hairy throughout its length. In some species, the horns resemble those of a Cape Buffalo. The animals are known for their hilarious behavior.

The Marsh-Antelopes appear and behave more like deer than horses, the qualities of both of the latter animals tending to blend in antelopes. The Waterbuck is a large antelope of southern and eastern Africa. It frequents rocky hills in the vicinity of rivers. The horns are more than 2 feet long, slightly curved and ringed almost to the tips. The Reedbuck is a comparatively small antelope. The male has small horns which turn forward. Also known by the Dutch equivalent, *reitbok*. The Rhebok also is small and is found in hilly sections of eastern and southern Africa. Some authorities have compared this animal with the chamois in its habits.

The Blackbuck, as indicated by the accompanying table, does not fit into the other large subfamilies of the antelopines. Of the antelopines, the blackbuck appears to be more closely related with the gazelles. The animal was known in Europe since the Middle Ages. Its present habitat is India and it prefers the open plains from the foothills of the Himalayas to Cape Comorin; and western Pakistan to lower Assam.

As indicated by their name, the Pigmy Antelopes are very small animals, probably the smallest of the ungulates (having hoofs). The Royal antelope of west Africa is only about 8 inches (20 centimeters) high. However, the other species of pigmy antelopes are considerably larger. The Klipspringer is a great jumper, sure-footed, and known for tripping about on its toes. The Oribis is of dainty build, standing less than 2 feet high (0.6 meter), with sharply pointed horns from 4 to 5 inches (10 to 13 centimeters) in length. The color is tawny above and white below. The Sunis is black with white rings around the eyes and white underparts and ears. The horns are long and thin.

The Gazelles comprise the largest group of antelopines. To all but zoologists and keen sportsmen, most gazelles appear very much alike. They are small, delicate, with preference for the desert, grassy plains, scrub zones, and parklands. The Impala is moderately large and of several species; most of these animals have long, slender horns, slightly spiraled and ringed through most of their length. The Gerenuk inhabits eastern Africa and has a very long neck and moderate-spiraled horns, turned forward sharply at the tips. Also called Waller's gazelle. The Addra or true gazelles are found all over Africa outside the forest zones, as well as in Syria, western Arabia, the plains of India, and central Asia from Turkey eastward to the Gobi desert of Mongolia. The affinity between the antelopines and the caprines is found in the so-called Gazelle-Goats. Among these are the Chiru, the Saiga, and the Goa. The latter animal prefers the high plateaus of Tibet. The gazelle-goats travel in small herds with a diminishing population because of hunting. The gazelle-goats are not to be confused with goat-gazelles which are described under **Goats and Sheep.**

ANTENNA (Communications). Characteristically, communication systems consist of cascaded networks, each network designed to carry out some operation on the energy conveying the information. In radio communication systems, antennas are the networks serving to transfer the signal energy from circuits to space and, conversely, from space to circuits. In circuits, the flow of energy is restricted to one or the other of two directions. The effectiveness of transfer of energy between the antenna and the adjacent circuit element is, therefore, determined solely by the terminal impedance of the antenna and that of the adjacent circuit. The knowledge of the antenna terminal impedance over the desired frequency range, therefore, fully describes the joint performance of the antenna and the circuit energy.

At one time, antennas were used mainly in connection with radio transmission and reception. In recent years, antennas have been developed for other portions of the radiation spectrum, including light waves.

The relationship between the antenna and space, however, is much more complex. The distribution of the radiated energy varies with the direction in space and with the distance from the antenna. This gives rise to the directive properties of the antenna. Further, the energy is radiated in the form of an electric and a magnetic field. These are vector quantities which, at a distance from the source, are at right angles to each other and to the direction of propagation. The planes in which these vectors are located, and whether they are stationary or rotate with time determine the polarization of the radiated field. The performance of an antenna can, therefore, be fully described only by specifying several parameters, such as radiation pattern, gain, and polarization.

It is convenient, in discussing antenna properties, to consider the antenna as a radiating rather than a receiving network. The antennas are, however, linear networks and are subject to the law of reciprocity. (As used here, radiation intensity has the dimensions of power flow per unit area, normally, watts per square meter. Electric field strength, on the other hand, is in volts per meter.)

The performance of an antenna in terms of radiation pattern, gain, or polarization is the same, irrespective of whether the antenna radiates or absorbs radiation.

Except for the immediate neighborhood of the antenna, referred to as the "near-field region" of the antenna, radiated energy propagates radially from the antenna, and the radiation intensity varies inversely as the square of the distance from the antenna. This is a propagation effect. In discussing the antenna performance, it is customary to disregard this and to represent the distribution of the radiated power as a function of the two direction angles only. Such a distribution is commonly represented graphically and is then known as the radiation pattern of the antenna. See Fig. 1.

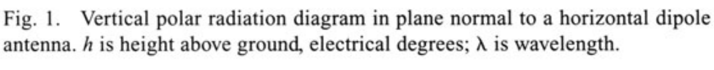


Fig. 1. Vertical polar radiation diagram in plane normal to a horizontal dipole antenna. *h* is height above ground, electrical degrees; λ is wavelength.

The radiation patterns can take a variety of forms. Sometimes they are in the form of a polar diagram, with the radial distance proportional to either field strength or intensity. The intensity may be represented linearly, as power, or logarithmically, in decibels. For representing the directive properties of an antenna in all directions, contours of equal radiation intensity may be plotted, with the two direction angles as abscissas and ordinates, respectively. See Fig. 2.

The directive properties of an antenna also lead to the concept of antenna "gain." The directive gain of an antenna in a specified direction is the radiation intensity in that direction compared to what it would be

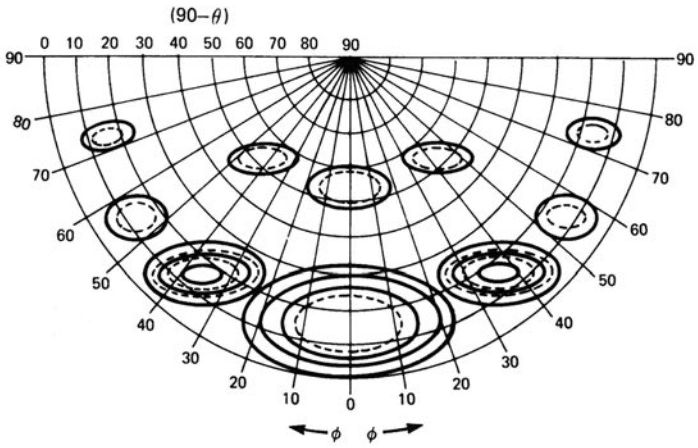


Fig. 2. Directive patterns for a rhombic antenna. Apex angle is 18°. Height above ground h is 0.8 λ (wavelength).

if the total radiated power were distributed equally in all directions. For some applications, such as point-to-point communication, high values of antenna gain are desired because such antennas concentrate the available power, thus effectively increasing it. Conversely, in receiving applications, such antennas are more reponsive to radiation arriving from one direction. For other applications, such as broadcasting, antennas with low directivity may be desired.

The gain of an antenna is dependent principally upon the size of the antenna, expressed in wavelengths. The larger the antenna, the greater is likely to be its gain. The values of gain for different antennas range from 1.5 for an electrically small dipole to hundreds and even thousands times that. In practice, antenna gains are usually expressed logarithmically, in decibels. For the low-frequency end of the radio spectrum (15 kHz to 3 MHz), antennas, although large physically, are relatively small in terms of wavelengths. Therefore, the directive gains of these antennas seldom exceed 3 (4.8 dB). In the high-frequency band (3 to 30 MHz), which is used principally for long-distance communication, antenna gains of 10 to 100 (10 to 20 dB) are frequently encountered. At microwave frequencies, where the wavelengths are a fraction of a meter, gains of several hundred, and even thousand times (20 to over 30 dB) are common.

When an antenna has one or more of its dimensions significantly larger than a wavelength, its radiation pattern is likely to have more than one maximum. The radiation pattern, in such cases, is said to have a lobe structure. That part of the radiation pattern which encompasses the direction of the largest maximum and the radiation immediately to each side of it, is referred to as the main lobe. The radiation about the minor maxima is referred to as the secondary or side lobes. One of the frequent goals in antenna design is the reduction in the levels of secondary lobes. These may, at times, be a source of interference to other transmissions.

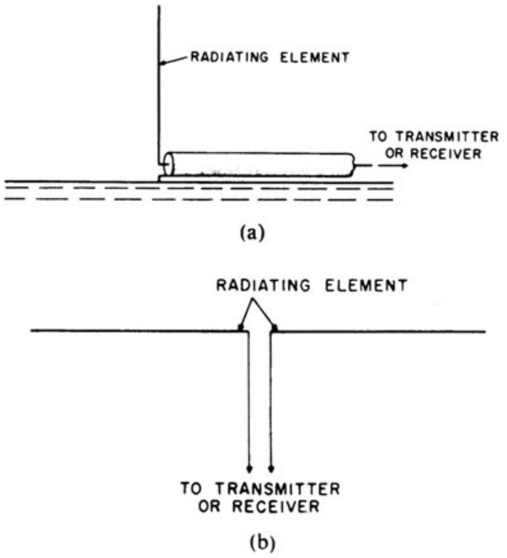
In common with light, radio waves consist of electric and magnetic fields at right angles to each other and to direction of propagation. In radio terminology, the orientation of the electric vector of a radio wave is taken as the direction of polarization. Thus, if the electric field vector is parallel to the ground, the radio wave is termed "horizontally polarized." Although the polarization of the energy radiated by an antenna, in general, varies with the direction, an antenna is usually designated as being horizontally (or vertically, circularly, etc.) polarized, depending on the polarization of its radiation in the direction of the main lobe maximum.

The importance of polarization in radio engineering lies principally in the different reflective properties of the ground for waves with electric field parallel to the ground and those normal to the ground. Different radio services are served best by different polarizations. Antennas for use in the low-frequency end of the radio spectrum, as previously defined, are almost invariably vertically polarized. This includes the AM broadcast band. In the high-frequency band, both horizontal and vertical polarizations are used. For FM and television broadcast service in the United States and many other countries (but not in the United Kingdom), horizontal polarization is employed.

Antenna Configurations

The types and variations of antennas are extremely numerous. Each type has some particular advantage over the others for some specific requirement. Among some of the more important and frequently encountered requirements are those for operating bandwidth, directivity, whether high or low, and polarization. Mounting factors for receiving antennas, as in land vehicles, aircraft, and spacecraft, also pose problems of size, weight, air resistance, etc.

A few of the representative types of antennas frequently encountered in practice are shown in Figs. 3, 4, and 5. Two very elementary radiators, a monopole and a dipole, are shown in Fig. 3. A monopole in one form or another is employed almost exclusively throughout the low-fre-



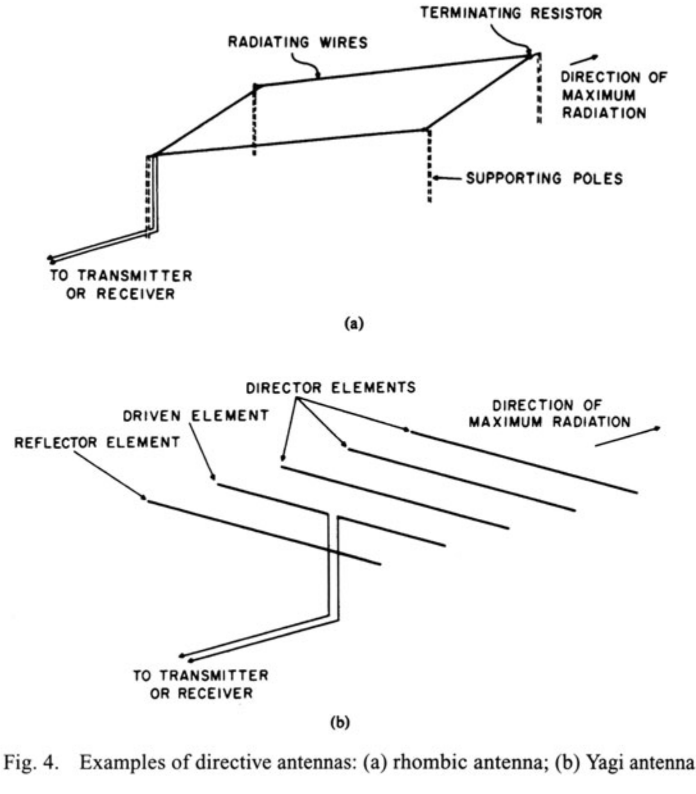


Fig. 4. Examples of directive antennas: (a) rhombic antenna; (b) Yagi antenna.

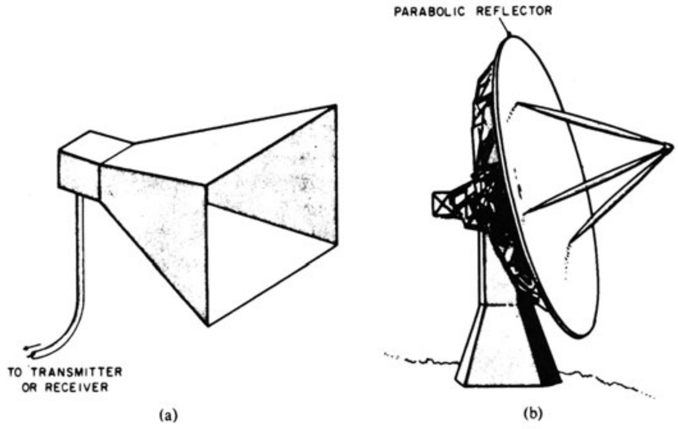


Fig. 5. Microwave-type antennas: (a) horn antenna; (b) parabolic-reflector antenna.

quency end of the radio spectrum. The dipole is somewhat more versatile, as it can be oriented to give either horizontal or vertical polarization. It is frequently used as an elementary radiator in large array-type antennas.

Figure 4 presents two highly directive, but otherwise radically different types of antennas. The rhombic antenna has broadband properties and is used widely in point-to-point communication service. The Yagi antenna is a relatively compact antenna with high gain for its size. Its operating frequency band is quite narrow. Two antennas used at microwave frequencies are shown in Fig. 5. The horn antenna is used generally where moderate directivity suffices. The parabolic antenna is used for high-gain applications and is a quasi-optical device.

Specific antenna configurations are described briefly in the following listing, arranged alphabetically for convenience.

(*Achromatic Antenna*)—An antenna whose characteristics are uniform over some band of frequencies.

(*Adcock Antenna*)—A form of radio antenna which in its simplest form (see Fig. 6) consists of two spaced vertical antennas. This type of antenna finds use in radio direction finding where it displays a significant advantage over the loop antenna. When the latter is used to obtain a bearing, incorrect results may be obtained because horizontally polarized downcoming radio waves will induce different voltages in the two horizontal members of the loop. In the Adcock antenna, on the other hand, horizontally polarized waves induce voltages in the horizontal elements which cancel in their effect on the output voltage.

(*Alford Slotted Tubular Antenna*)—A horizontally polarized antenna developed for FM broadcast work. It consists of a sheet of metal bent into the form of a cylinder which is not quite closed, hence a straight narrow slot extends the full length of the cylinder or tube. It is so dimensioned that the distribution of potential across the slot has very nearly the same phase throughout the entire length of the slot. The currents produced flow in horizontal circles around the cylinder so that the latter operates something like a stack of small, in-phase loops.

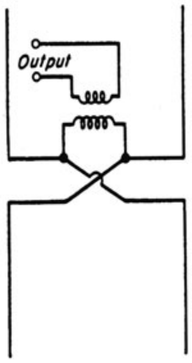


Fig. 6. Elementary Adcock antenna.

(*Aperiodic Antenna*)—A nonresonant, and thus frequently insensitive antenna.

(*Base-Loaded Antenna*)—An antenna (usually vertical) whose electrical height is increased by the addition of inductance in series with the antenna at the base.

(*Biconical Antenna*)—An antenna formed by two conical conductors, having a common axis and vertex, and excited at the vertex. When the vertex angle of one of the cones is 180°, the antenna is called a discone.

(*Broadband Antenna*)—An antenna which will function satisfactorily over a bandwidth in the order of 10% or more of its center frequency.

(*Cage Antenna*)—An antenna in which the radiating members are parallel rods arranged in a cylindrical fashion.

(*Capacitor Antenna*)—An antenna in which the capacitance between two conductors or systems of conductors is the essential characteristic. Also called a dielectric antenna.

(*Cheese Antenna*)—A cylindrical parabolic reflector enclosed by two plates perpendicular to the cylinder, so spaced as to permit the propagation of more than one mode in the desired direction of polarization. It is fed on the focal line.

(*Cloverleaf Antenna*)—An antenna for transmission or reception of horizontally polarized radiation in a nondirectional pattern in a plane normal to the axis of the antenna. Its name arises from the fact that it is comprised of a cluster of four half-wave, curved, radiating elements arranged in the pattern of a four-leaf clover.

(*Coaxial Antenna*)—An antenna comprised of a quarter wavelength extension to the inner conductor of a coaxial line, and a radiating sleeve which, in effect, is formed by folding back the outer conductor of the coaxial line for approximately one-quarter wavelength.

(*Conical Antenna*)—A wideband antenna in which the driven element or elements are conical in shape.

(*Cosecant-Squared Antenna*)—A type of radar antenna which provides constant field strength at a given altitude over an appreciable range. So-called because the power in the antenna pattern on one-way transmission decreases at a rate proportional to the square of the cosecant of the elevation angle. This radiation pattern may be approximated by using a paraboloidal reflector with several radiators arranged in a line perpendicular to the axis.

(*Dielectric Antenna*)—An antenna which employs a dielectric as the major component in producing the required radiation pattern.

(*Dipole Antenna*)—A straight radiator, usually fed in the center, and producing a maximum of radiation in the plane normal to its axis. The length specified is the overall length. Common usage in microwave antennas considers a dipole to be a metal radiating structure which supports a line current distribution similar to that of a thin straight wire, a half-wavelength long, so energized that the current has two nodes, one at each of the far ends. See Fig. 3(b).

(*Directional Antenna*)—For the usual radio-broadcast service, it is desirable to transmit the signal in all directions equally, but for special broadcast services, such as international short-wave or microwave relay transmission, it is often desirable to direct the radiation in some specific direction and avoid radiation in other directions. The need for directed radiation is even more pronounced in other types of radio service. The radio signals may be directed by the use of directional antennae, or, as often called when consisting of more than one element, directional arrays. Any antenna is directional to a certain extent, e.g., the common tower antenna for broadcast stations does not radiate directly upward, but in the sense used here a directional antenna is one having marked characteristics of this type. Basically the directional antennae all depend upon radiation from two or more components adding vectorially. If the waves radiated from various elements add in a certain direction the signal will be strong in that direction, while if they tend to cancel, or subtract, in a given direction the signal will be zero or weaker in that direction.

One of the simplest directional antennas is the loop such as used with many portable receivers. Here the two elements whose effects add vectorially are the two vertical sides of the loop. The result is a figure 8 radiation pattern, i.e., if lines are drawn to scale in various directions so their lengths represent the strength of the signal in each direction, the ends will all lie on a figure 8 curve with the antenna at the center. These antennas are used for many radio ranges, and, since the directional characteristics of any antenna are the same for transmission and reception, for radio compass use. The directional pattern may be altered by adding the radiation from a separate vertical antenna.

For broadcast use where it is necessary to decrease the signal in certain directions, usually to avoid interference with another station, systems consisting of two or more vertical antennas are quite common. By proper spacing of the elements and proper choice of the phase of the currents (which can easily be adjusted by circuit values) a wide range of radiation patterns may be obtained.

For international short-wave broadcasts and for point-to-point communication, more elaborate extensions of the same principle are used. Since the more elements in an array, the sharper the pattern, the radiation may be beamed at will, the type of service and economics being the usual limiting factors. By stacking systems one above the other in a vertical plane the radiation may be directed vertically as well as horizontally. It should be mentioned that it is not desirable to have the beam too sharp even for point-to-point service since the variations in the ionosphere may cause the signal to miss the receiver if the beam is too sharp.

Sometimes the elements of an array are not all fed directly from the transmitter, but some are fed and others pick up energy radiated by the first and reradiate it. By proper choice of the spacing and the antenna dimensions these various radiations may be made to give the desired pattern. The fed antennas are often referred to as driven antennas or elements and the others as parasitic antennas. There are many other types of directional antennas, such as rhombic, V, herringbone, binomial, broadside, continuous-spaced, linear, etc., but all depend upon vector addition of the radiation to give the pattern.

By the use of directional arrays the signal transmitted in a given direction may be increased manyfold over its value for the same transmitter with a nondirectional antenna. A measure of this is the gain of the array which is the signal with the array divided by the signal in the same direction for one element of the array serving as antenna. Where the type service permits their use, directional arrays are the most economical means of obtaining increased signal strength at the receiver. In reception the directional antenna allows the reception of a signal from the desired direction and suppresses signals and noise from other directions.

(*Discone Antenna*)—An antenna of a disk and a cone whose apex approaches and becomes common with the outer conductor of the coaxial feed at its extremity. The center conductor terminates at the center of the disk, which is perpendicular to the axis of the cone. Its most important characteristic is its ability to operate over a very wide bandwidth without a substantial change of input impedance or radiation pattern. The radiation pattern is omnidirectional in a plane perpendicular to the axis of the cone.

(*Dummy Antenna*)—A substitute for an actual antenna which is used for test purposes. In making comparative tests, calibrations, etc., on receivers, it is highly desirable to have conditions as near as possible to actual used conditions, yet have them standardized so they may be reproduced or the results on different units accurately compared. To do this a standard dummy antenna is used. The make-up of the antenna varies with different types of sets, being a series circuit with an inductance of 20 microhenries, capacitance of 200 micromicrofarads and resistance of 25 ohms for regular broadcast receivers. For auto radios, short-wave sets, etc., other circuit combinations are standard. The dummy antenna is connected between the set and the standard signal generator which supplies the radio-frequency test voltages. A dummy antenna consisting of just resistance is frequently used as a load on radio transmitters for making preliminary adjustments without radiating a signal. The output power of the transmitter is dissipated as heat in the resistance.

(*Fanned-Beam Antenna*)—A unidirectional antenna so designed that transverse cross sections of the major lobe are approximately elliptical.

(*Fishbone Antenna*)—An antenna consisting of a series of coplanar elements arranged in colinear pairs, loosely coupled to a balanced transmission line.

(*Franklin Antenna*)—An antenna consisting of a number of half-wave dipoles placed end-to-end, all operating in phase. Also called a colinear antenna.

(*Helical Antenna*)—An antenna used where circular polarization is required. The driven element consists of a helix supported above a ground plane. If the circumference of one turn is approximately one-half wavelength, the radiation is said to be in the axial mode and is directed predominately along the axis of the helix. In this mode, the antenna has good efficiency and relatively broad bandwidth.

(*Horn Antenna*)—A circular or rectangular electromagnetic horn can be used for providing unidirectional pattern coverage with either linear or circular polarization. Circular polarization can be achieved by exciting a square waveguide with a signal polarized at 45° and placing a quarter-wave plate between the feed point and the aperture. The waveguide and transition feeding it essentially determine the bandwidth of a horn. The bandwidth can be increased by using ridge waveguides and other broadbanding techniques. See Fig. 5(a).

(*Image Antenna*)—An antenna located close to the earth's surface (assumed to be a perfectly conducting plane) transmits a direct ray and a ray reflected from the earth's surface. It is convenient to represent the reflected ray as originating from an image antenna identical to the original, and located inside the earth by a distance equal to the height of the original above the earth.

(*Isotropic Antenna*)—Sometimes referred to as a unipole, this is a hypothetical antenna radiating or receiving equally in all directions. A pulsating sphere is a unipole for sound waves. In the case of electromagnetic waves, unipoles do not exist physically, but represent convenient reference antennas for expressing directive properties of actual antennas.

(*J Antenna*)—A half-wave antenna, end-fed by a parallel-wire, quarter-wave section having the configuration of a letter J.

(*Lazy H Antenna*)—An antenna array where two or more dipoles are stacked one above the other for the purpose of obtaining greater directivity.

(*Leaky-Pipe Antenna*)—External radiation is produced by providing a hole or slot in a waveguide propagating electromagnetic power. Proper choice of the size and location of a series of holes in the waveguide may lead to quite directional radiation patterns.

(*Lens Antenna*)—To satisfy high directivity requirements, a lens is often placed in front of another radiator, such as a dipole or horn. In much the same manner as an optical lens focuses light waves, these microwave lenses focus the high-frequency energy into a sharp beam.

(*Logarithmic Antenna*)—In logarithmic antennas, the radiating elements are in geometric progression in accordance with their resonant frequencies. This results in effectiveness at a wide range of frequencies. Sometimes used for special purposes in the high-frequency field. The radiating elements are folded dipoles. The aperture efficiency is comparatively low because only those elements that are fairly close to resonance will contribute to the radiation.

(*Log-Periodic Antenna*)—The geometry of this antenna repeats periodically so that the electrical properties of the antenna also repeat periodically with the logarithm of the frequency. These antennas are essentially frequency-independent and capable of bandwidths of 10:1 or greater, and with little change in patterns.

(*Loop Antenna*)—A loop antenna, when used with radio transmitters or receivers, possesses valuable directional properties.

On ships and airplanes, a loop antenna equipped with a compass card, whose 000°–180° line is parallel to the longitudinal axis of the ship, may be used to obtain relative bearings of radio stations. In using the loop, the position of minimum intensity is sought. In other words, the indicating needle on the compass card is in a direction perpendicular to the plane of the loop, and the observer rotates the antenna until the position of minimum signal intensity is found. The reading of the dial will then give a line of position through the ship and the radio station. With the simple loop there is no method for telling on which side of the instrument the station is located. To overcome this difficulty, the loop antenna is combined with a nondirectional antenna.

(*Loop-Vee® Antenna*)—A broadband antenna that produces a circularly polarized pattern. Pattern coverage of the grounded Loop-Vee is practically identical to that produced by a quarter-wave stub in conjunction with a current loop above the ground plane. The azimuth patterns are nearly perfect circles for all polarizations. In another design, the balanced Loop-Vee incorporates a second element similar to the single element of a grounded antenna and produces a butterfly-shaped pattern.

(*Marconi Antenna*)—An antenna that has one end of its radiating surface grounded (a grounded antenna). It may be considered a transmission line open-circuited at the far end and driven at the sending end, which is the junction between antenna and ground.

(*Marconi-Franklin Antenna*)—This array, one of the first used for high-speed short-wave point-to-point communication, consists of a front curtain of vertical radiators, each consisting of several cophased dipoles in series, and another curtain of reflecting wires of the same construction situated one-quarter wavelength to the rear. There are twice as many reflectors as radiators.

(*Monopole Antenna*)—Any one of several configurations of very simple design, often applied to vehicles, which protrude vertically from the vehicle, such as a whip, spike, blade, cone, or sleeve. See Fig. 3(a).

(*Multiple-Tuned Antenna*)—A low-frequency antenna having a horizontal section with a multiplicity of tuned vertical sections.

(*Musa Antenna*)—A "multiple-unit steerable antenna" consisting of a number of stationary antennas, the composite major lobe of which is electrically steerable.

(*Omnidirectional Antenna*)—An antenna producing essentially constant field strength in azimuth, and a directive radiation pattern in elevation.

(*Oscillating Doublet Antenna*)—A reference standard against which the directional characteristics of an antenna may be compared. Ideally it consists of two closely spaced charges of opposite sign, both oscillating in the same phase. Also, it may be regarded as an infinitely short, linear current-element.

(*Parabolic Antenna*)—A directional antenna using some form of a paraboloidal mirror either to convert plane waves into spherical waves or to convert spherical waves into plane waves. The reflector is fed or "illuminated" by the use of dipoles, waveguide feed system, or horns. The simple parabolic mirror is truly a broadband device. See Fig. 5(b).

(*Pencil-Beam Antenna*)—A unidirectional antenna so designed that cross sections of the major lobe by planes perpendicular to the direction of maximum radiation are approximately circular.

(*Pill-Box Antenna*)—A cylindrical, parabolic reflector enclosed by two plates perpendicular to the cylinder, so spaced as to permit the propagation of only one mode in the desired direction of polarization. It is fed on the focal line.

(*Pocket Antenna*)—A nonprotruding slot antenna developed for aircraft.

(*Quarter-Wave Antenna*)—An antenna which is electrically one-quarter of a wavelength long. It may be physically longer or shorter than one-quarter wavelength in free space.

(*Rhombic Antenna*)—An antenna composed of long-wire radiators comprising the sides of a rhombus. The antenna usually is terminated in an impedance. The sides of the rhombus, the angle between the sides, the elevation, and the termination are proportioned to give the desired directivity. See Fig. 4(a).

(*Scanning Antenna*)—A directional antenna employed in radar which mechanically or electrically causes its radiation to periodically scan a given arc or solid angle.

(*Series-Fed Vertical Antenna*)—A vertical antenna which is insulated from ground and energized at the base.

(*Shaped-Beam (Phase-Shaped) Antenna*)—A unidirectional antenna whose major lobe differs materially from that obtainable from an aperture of uniform phase. A $\operatorname{cosec}^2 \theta$ beam is a shaped beam whose intensity in some plane varies as $\operatorname{cosec}^2 \theta$ over a prescribed range, where θ is a polar angle in that plane. The half-power width in planes perpendicular to this plane is approximately constant for the prescribed range of θ.

(*Shunt-Fed Vertical Antenna*)—A vertical antenna connected to ground at the base and energized at a point suitably positioned above the grounding point.

(*Slot Antenna*)—A radiating element formed by a slot in a metal surface.

(*Steerable Antenna*)—A directional antenna whose major lobe can be readily shifted in direction.

(*Top-Loaded Vertical Antenna*)—A vertical antenna so constructed that, because of its greater size at the top, there results a modified current distribution giving a more desirable radiation pattern in the vertical plane. A series reactor may be connected between the enlarged portion of the antenna and the remaining structure.

(*Tridipole Antenna*)—An omnidirectional, horizontally polarized antenna consisting of three dipoles displaced from each other by 60° in the horizontal plane. The radiators are curved, causing the array to have a circular appearance.

(*Turnstile Antenna*)—An antenna composed of two dipole antennas, normal to each other, with their axes intersecting at their midpoints. Usually, the currents are equal and in phase quadrature.

(*V Antenna*)—A V-shaped arrangement of conductors, balanced-fed at the apex, and with included angle, length, and elevation proportioned to give the desired directivity.

(*Yagi Antenna*)—An array with one or more parasitic elements in addition to the driven element or elements. Currents induced in the parasitic element from the field produced by the driven antenna cause radiation in a phase (relative to the phase of the radiation from the driven unit) that is a function of both the spacing of the elements and the length of the parasitic element. If the resultant radiation pattern has its maximum in the direction of the driven element, the parasitic element is called a reflector, whereas if the maximum radiation is in the direction of the parasitic antenna, it is called a director. See Fig. 4(b).

Special Antennas

An antenna custom-designed for use with the *large radio telescope* at the Arecibo Observatory (Arecibo, Puerto Rico) is shown in Fig. 7. As illustrated further in the entry on **Radio Astronomy,** a triangular platform measuring 216 feet (66 meters) on each side, is 500 feet (152 meters) in the air and above a 1000-foot (305-meter) diameter reflector

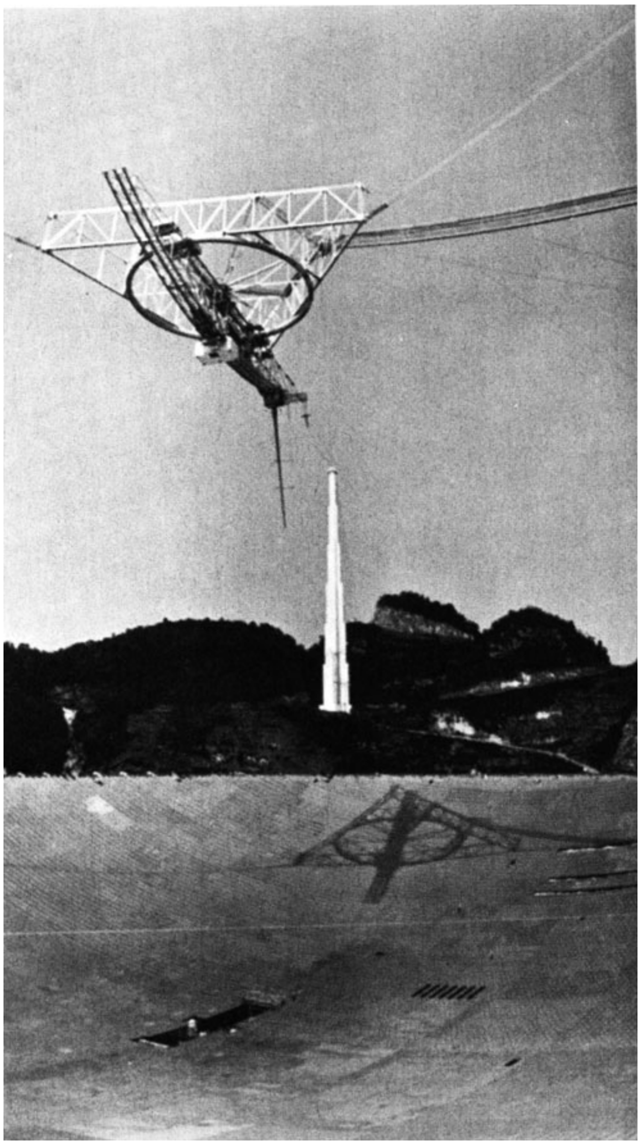


Fig. 7. Special antenna for use with large radio telescope (Arecibo, Puerto Rico). (*The National Astronomy and Ionosphere Center*, Cornell University.) (*Photo by Russell C. Hamilton.*)

bowl. The output of the transmitter is 450,000 watts. This power is maximized by the positioning of the transmitter in a carriage house of the suspended platform, instead of on the ground. At the short wavelengths (S-Band), much power would be lost if the signal had to travel a considerable distance from the ground to the platform even through the most efficient S-band waveguide obtainable. Thus, the decision to put the transmitter on a suspended platform was made. The transmitter had to be extremely compact. The transmitter, despite its high efficiency gives off as much heat energy as it does radio energy. Heat exchangers were built on the platform to dissipate this energy, with coolant water circulating between the platform and the carriage house through a long, articulated piping system.

A special antenna used for high-frequency satellite communications studies and radio astronomy research is shown in Fig. 8.

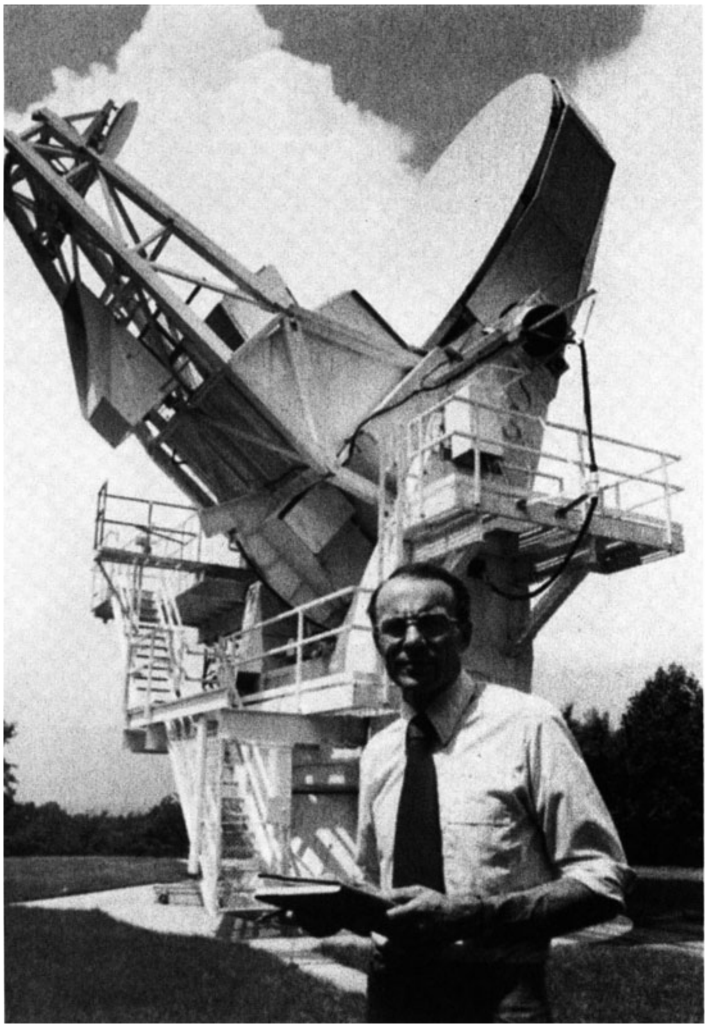


Fig. 8. Special antenna for high-frequency satellite communications studies and radio astronomy research. (*AT&T Bell Laboratories.*)

Additional Reading

Benson, K. B., and J. Whitaker: "Television Engineering Handbook," McGraw-Hill, New York, 1992.

Bollthias, L.: "Radiowave Propagation," McGraw-Hill, New York, 1988.

Carison, A. B.: "Communication Systems," 3rd Edition, McGraw-Hill, New York, 1986.

Dane, A.: "Arrays Turn Aircraft Skins Into Radar Antennas," *Popular Mechanics*, 15 (January 1990).

Dayton, R.: "Telecommunications," McGraw-Hill, New York, 1991.

Fink, D. G., and D. Christiansen: "Electronics Engineers' Handbook," 3rd Edition, McGraw-Hill, New York, 1989.

Kaufman, M., and A. H. Seldman: "Handbook of Electronics Calculations for Engineers and Technicians," 2nd Edition, McGraw-Hill, New York, 1988.

Kraus, J. D.: "Antennas," 2nd Edition, McGraw-Hill, New York, 1988.

Lee, W. C. Y.: "Mobile Cellular Telecommunications Systems," *McGraw*-Hill, New York, 1989.

Maillox, R. J.: "Microwave and mm-Wave Array Antennas," *Microwave J.*, 17 (January 1990).

Rohde, U., and T. T. Bucher: "Communications Receivers: Principles and Design," McGraw-Hill, New York, 1988.

Seeds, A.: "Optical Beamforming Techniques for Phased-Array Antennas," *Microwave J.*, 14 (July 1992).

Staff: "Multi-Function Antenna Uses Optical Signals," *HughesNews*, 1 (June 1, 1990).

Staff: "A Circular Polarized, Low Profile Antenna," *Microwave J.* 113 (April 1992).

Steyskal, H.: "Array Error Effects in Adaptive Beamforming," *Microwave J.*, 101 (September 1991).

Wiltse, J. C., and J. E. Garrett: "The Fresnel Zone Plate Antenna," *Microwave J.*, 101 (January 1991).

ANTENNA (Zoology). A jointed sensory appendage of the head found in several classes of *Arthropoda*. Crustacea have two pairs, while insects, centipedes, and millipeds have one pair. See also **Diplopoda.**

ANTHER. The terminal part of a stamen, containing the pollen sacs. See **Flower.**

ANTHERIDIUM. The structure which gives rise to the sperm. In the algae it is a single cell, the contents of which may become a single sperm or divide to produce many sperms. In the higher divisions of plants, the antheridium is a multicellular body which contains the sperms.

ANTHESIS. That stage in the flowering development of a plant when pollen is being produced.

ANTHOCYANINS. A group of water-soluble pigments which account for many of the red, pink, purple, and blue colors found in higher plants. Most plants contain more than one of these pigments and they occur most prevalently as glycosides. Several hundred different anthocyanins are known. Anthocyanins have been isolated and some have been found to be acylated with substituted cinnamic acids. The site of attachment of these acids to the anthocyanins has not been fully definitized. The natural role of the anthocyanins in plants to date has not been related to any factor of plant metabolism and many authorities believe that the pigments play more of an ecological role in regard to pollination and seed dispersal through their ability to act as an insect and bird attractant.

The anthocyanins are part of the larger group of aromatic oxygen-containing, heterocyclic compounds, known as flavonoids, most of which have a 2-phenylbenzopyran skeleton as their basic ring system. Although widely distributed among higher plants, including ferns and mosses, they are not found in algae, fungi, bacteria, or lichens.

There has been considerable interest and research activity in connection with anthocyanins during the past decade or so, stemming principally from the tighter restrictions, including banning, of several synthetic colorants. See also **Colorants.** Representative of the food processing industry's desire to find colorants that are beyond suspicion as health deterrents, scientists have been investigating various sources of anthocyanins. They have found that pigments from roselle plants (*Hibiscus sabdariffa*) native to the West Indies can be used for coloring apple and pectin jellies. A cranberry pomace extract has been found useful in coloring cherry pie filling. The potential of blueberry as a source of anthocyanin pigments also has been investigated. The berry is rich in nonacylated anthocyanins, but presently appears to be too costly as a coloring substitute.

Grape anthocyanins have been intensely investigated and have been found reasonably satisfactory, for example, in carbonated beverages. Although to date the grape anthocyanins are not as stable as Red No. 2, research continues, encouraged by the large amounts of grape wastes produced in the production of wine and grape juice. Red cabbage also has been seriously considered as a source of anthocyanin pigments. Because the anthocyanins are most stable at a pH range of 1.0 to 4.0, this acidity dictates the products in which they can be used.

Much more detail on this topic can be found in the references listed below. See also **Colorants (Foods); Glycosides;** and **Pigmentation (Plants).**

Additional Reading

Ballinger, W. E., Maness, E. P., and L. J. Kushman: "Anthocyanins in Ripe Fruit of the Highbush Blueberry (*Vaccinium corymbosum* L.)," *J. Amer. Soc. Horticultural Sci.*, **95**, 283 (1970).

Clydesdale, F. M., et al.: "Concord Grape Pigments as Colorants for Beverages and Gelatin Desserts," *J. Food Sci.*, **43**, 6, 1687–1692 (1978).

Considine, D. M. (editor): "Foods and Food Production Encyclopedia," Van Nostrand Reinhold, New York, 1982.

Newsome, R. L.: "Food Colors—Scientific Status Summary," *Food Technology*, **40**(7), 49–56 (July 1986).

Shewfelt, R. L., and E. M. Ahmed: "Anthocyanin Extracted from Red Cabbage Shows Promise as Coloring for Dry Beverage Mixes," *Food Product Development*, **11**, 4, 52–58 (1977).

Volpe, T.: "Cranberry Juice Concentrate as a Red Food Coloring," *Food Product Development*, **10**, 9, 13 (1976).

ANTHOPHYLLITE. The mineral anthophyllite is an orthorhombic amphibole essentially (Mg, Fe)$_7$Si$_8$O$_{22}$(OH)$_2$ with aluminum sometimes present. This mineral corresponds to enstatite and hypersthene in the pyroxene group. It has a prismatic cleavage; hardness, 5.5–6; sp gr, 2.8–3.57; luster, vitreous; color, gray, yellow, brown, green or brownish-green; transparent to translucent; probably always a metamorphic mineral in magnesium-rich rocks, often associated with talc; very common in schists. Found in Norway, Austria, Greenland, Pennsylvania, Georgia, and elsewhere. The name is derived from the Latin *anthophyllum*, clove, because of its usual brownish shades. See also **Amphibole.**

ANTHOZOA. The sea anemones, corals, alcyonarians and related forms. A class of the phylum *Coelenterata* in which the polyp form gains its highest development and the medusa is unknown.

Like the hydrozoan polyps, these animals have relatively thin walls, due to the thin middle layer (mesogloea), and are approximately cylindrical in form. The base is a disk by which the animal is attached to some support and the opposite end forms an oral disk bearing numerous hollow tentacles surrounding the mouth. The mouth leads into a long tube lined with ectoderm, known as the stomodaeum. In it ciliated grooves serve for the passage of currents of water into and out of the enteric cavity. In this cavity radiating partitions, the mesenteries, pass from the wall to the stomodaeum, which they hold in place. Others extend into the cavity from the wall without reaching the stomodaeum. The edges of the mesenteries bear mesenteric filaments with stinging cells. They are important in digestion and respiration. Reproductive bodies also develop in the mesenteries and slender acontia with many stinging cells arise from their edges. Muscle bands in the mesenteries and in the body wall contract the entire animal and close the margins of the oral disk in over the tentacles.

The class is divided into two subclasses:

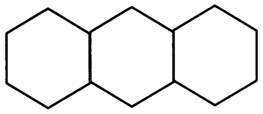
Subclass *Alcyonaria*. Polyp with eight tentacles, pinnately branched. Colonial forms, usually supported by a hard skeleton. The sea fans, precious coral, and sea feathers.

Subclass *Zoantharia*. Colonial or solitary. Polyp with few to many tentacles, not pinnately branched. Hard deposits formed under the basal disk in some species. The stony corals and sea anemones.

ANTHRACENE. A colorless solid; melting point 218°C, blue fluorescence when pure; insoluble in water, slightly soluble in alcohol or ether, soluble in hot benzene, slightly soluble in cold benzene; transformed by sunlight into para-anthracene ($(C_{14}H_{10})_2$).

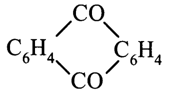
Anthracene reacts: (1) With oxidizing agents, e.g., sodium dichromate plus sulfuric acid, to form anthraquinone, $C_6H_4(CO)_2C_6H$. (2) With chlorine in water or in dilute acetic acid below 250°C to form anthraquinol and anthraquinone, at higher temperatures 9,10-dichloroanthracene. The reaction varies with the temperature and with the solvent used. The reaction has been studied using, as solvent, benzene, chloroform, alcohol, carbon disulfide, ether, glacial acetic acid, and also without solvent by heating. Bromine reacts similarly to chlorine. (3) With concentrated sulfuric acid to form various anthracene sulfonic acids. (4) With nitric acid, to form nitroanthracenes and anthraquinone. (5) With picric acid (1)HO · $C_6H_2(NO_2)_3$(2,4,6) to form red crystalline anthracene picrate, melting point 138°C.

$C_{14}H_{10}$ or



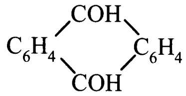
Anthracene is obtained from coal tar in the fraction distilling between 300° and 400°C. This fraction contains 5–10% anthracene, from which, by fractional crystallization followed by crystallization from solvents, such as oleic acid, and washing with such solvents as pyridine, relatively pure anthracene is obtained. It may be detected by the formation of a blue-violet coloration on fusion with mellitic acid. Anthracene derivatives, especially anthraquinone, are important in dye chemistry.

ANTHRAQUINONE. Anthraquinone (9,10) is a yellow solid, melting point 286°C; can be sublimed;

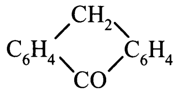


forms monoxime, melting point 224°C, by heating under pressure at 180°C with hydroxylamine chloride; forms no phenylhydrazone with phenylhydrazine; with strong oxidizing agents reacts with difficulty to yield phthalic acid $C_6H_4(COOH)_2$(1,2); with reducing agents, such as sodium hyposulfite, zinc in sodium hydroxide solution, tin or stannous chloride in hydrochloric acid (but not sulfurous acid), is reduced to anthraquinol, anthrone, dianthrol and dianthrone, depending on the conditions.

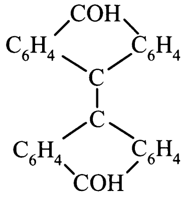
anthraquinol



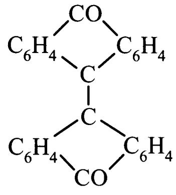
anthrone



dianthrol



dianthrone



Anthraquinone is obtained by oxidation of anthracene using sodium dichromate plus sulfuric acid, and is purified by dissolving in concentrated sulfuric acid at 130°C and pouring into boiling water, whereupon anthraquinone separates as pure solid, and is recovered by filtration. Further purification may be accomplished by sublimation or crystallization from nitrobenzene, aniline or tetrachloroethane. Anthraquinone is used as the material from which many dyes are made, notably alizarin $C_6H_4(CO)_2C_6H_2(OH)_2$ and related substances. These are vat dyes, that is, insoluble colored substances which are readily reduced to a substance having marked affinity for the fiber to be dyed and which upon exposure to the air are readily reoxidized to the original dye. Anthraquinone may be detected by the appearance of a red color on treatment with alkali, zinc powder, and water. See also **Coal Tar and Derivatives.**

ANTHRAX. This is a highly infectious disease caused by the gram-positive *Bacillus anthracis*. The disease is of historical importance because the anthrax bacillus was the first microorganism proved definitely to be the cause of an infectious disease. Anthrax mainly is a serious disease of cattle. It is capable of transmission to humans by way of meat and animal products. In a number of countries, extensive cattle vaccination programs have been very effective. These programs have virtually eliminated reservoirs of infection in the United States, where only minor outbreaks involving very few cases may occur, usually in the Great Plains, the lower Mississippi Valley, and Texas. Infection of humans usually occurs in persons who work with animal products, such as hides. In other parts of the world, anthrax in animals persists as a major problem, notably in Asia, Africa, and South America. The Russian literature also has reported outbreaks over many years. Outbreaks were particularly severe in 1923 and 1940, but the most disastrous of all occurred in and around the city of Sverdlovsk in 1979. Over 1000 fatalities were reported in an epidemic that lasted about one month. Even though the soil of Sverdlovsk province has been known to be infected with anthrax for a century or more, the characteristics and extent of the 1979 outbreak have baffled scientists. Rather than the mild cutaneous form, or the intermediately serious intestinal form, sketchy reports from Sverdlovsk indicated that it was the inhalation form of the disease that proved so potent. Questions which remain unresolved were immediately raised concerning the possibility of anthrax spores being released to the atmosphere as the result of an accident at a biological warfare facility.

The pathogenicity of *B. anthracis* stems from both the nature of its capsule and from toxin production. The former establishes the infection and the latter contributes to the characteristic edema around an anthrax lesion and to toxemia.

The most common form of anthrax is *cutaneous*. The cutaneous lesion is sometimes referred to as a malignant pustule. The bacillus forms spores in the external environment, and in culture, but not in animal tissues. Spores can linger in the soil and in animal products for many years. A cutaneous infection is usually the result of introducing spores through an abrasion. The cutaneous form can be successfully treated with antibiotics.

The intestinal form results from ingesting anthrax-contaminated meat, notably sausage. Inhalation anthrax, previously known as Wool-Sorter's disease, gained prominence as the result of the Sverdlovsk incident. The disease begins with a high fever, malaise, and non-productive cough. Dyspnea, cyanosis, hemoptysis, and chest pain follow. A striking feature is an extensive hemorrhagic mediastinitis. Before antibiotics were available, mortality from cutaneous anthrax ranged up to 20%. The mortality rate for untreated inhalation anthrax and anthrax meningitis is nearly 100%.

In the early part of this century, prior to the availability of antibiotics and the initiation of cattle vaccination programs, there were about 125 cases per year of cutaneous anthrax in the United States. For several years, the number of cases per year has been 10 or fewer.

Penicillin over a week is the treatment of choice, with tetracycline as an alternative for patients allergic to penicillin. Anthrax lesions should not be incised and drained because pus is usually absent and infection may be disseminated by this procedure.

R. C. V.

ANTHRAXOLITE. A coal-like metamorphosed bitumen, often closely associated with igneous rocks. Commonly associated with "Herkimer Diamond" type quartz crystals in dolomitic limestones in Herkimer and Montgomery counties in New York State.

ANTHROPOGENIC. Relative to the activities of humans as contrasted with the actions of natural forces and events. Thus, the addition of dust to the atmosphere from aerial crop spraying is anthropogenic, as contrasted with dust clouds produced by a volcano (natural); the pollution arising from an oil spill is anthropogenic, versus the seepage of crude oil into the oceans from cracks in undersea rock formations (natural).

ANTHROPOIDS (*Mammalia, Primates*). This division of *Mammalia* includes a number of medium-size and large animals which have comparatively large brains and which have no tails. Zoologically, *Hominidae* are also placed in this category. The subdivisions of apes include:

- Lesser Apes (*Hylobatidae*)
  - The Siamang (*Symphalangus*)
  - Gibbons (*Hylobates*)
- Greater Apes (*Pongidae*)
  - Gorillas (*Gorilla*)
  - The Chimpanzees (*Pan*)
  - Orangutans (*Pongo*)

The lesser and greater apes are found in the forests of Africa and in the Oriental Region (eastern India to Hainan and southward to Borneo, Java, and Sumatra).

There are about six species of gibbons, of which one, the siamang, is confined to Sumatra. Of the lesser apes, the siamang is the largest, having a finger-tip-to-finger-tip spread across the chest of nearly 6 feet (1.8 meters) and a height of about 3 feet (1 meter) when sitting in an upright position. Siamangs prefer high forest country and only infrequently travel on the ground where they do so in a rather awkward manner. When required, they are good swimmers, keeping their heads well out of the water. They are known for their frightening howling which occurs at dawn and sundown. Their preferred diet is fruit. They have a coat of long, rather shaggy black hair with the exception of gray beards in the males. These animals are characterized by a pouch connected to the throat which can be inflated to appear as a rather large red balloon.

Gibbons are smaller and of more slender build than the other apes and are among the most agile of all primates in the trees. See Fig. 1. They have extremely long arms with which they swing distances up to 20 feet (6 meters) from bough to bough. Their movements are quite rapid. This arm-swinging process is known as *brachiation*. Gibbons, as do spider monkeys, have features, in addition to their long limbs, which further the effectiveness of brachiation. including: The bone structure of the hand is formed in what might be termed a "hook" or "grapple"; a reduction of the thumb, such that it does not interfere with rapid release when going from one tree limb to the next; a curling-inward position of the hand when at rest; and exceptional elongation of the fingers and palm of the hand to provide increased surface area for contacting a tree limb as a leap is being completed. Some authorities consider the tree-swinging agility of the gibbon as unsurpassed.

Fig. 1. Mother gibbon and baby. (*A. M. Winchester.*)

A feature not found among other anthropoid apes is the presence of ischial callosities (small, naked, hardened, and thickened places) on the buttocks. For their diet, gibbons prefer leaves and fruits. The whooping noises of the siamang are also made by the gibbons, particularly for a few hours after dawn. Gibbons are well known for their intelligence, cleanliness, and gentility. Gibbons are monogamous and known for their exceptional fidelity and family cooperation. B. B. Beck (*Science*, **182**, 4112, 594–596 (1973)) reports that a bonded pair of hamadryas baboons developed use of a cooperative tool without training. The male could get food with the tool, but first had to get the tool from an adjoining cage which he could not enter. The female learned to give him the tool.

Specific variations of gibbons include: Hoolock (central Himalayas—gray to brown in color); White-handed Gibbon (Burma, Malay Peninsula—brown to black); Agile Gibbon (Sulu Archipelago and Borneo—cream-to-dark-brown color); Unkaputi (Malaysia—often taken as pets); and the Wow-Wow or Silvery Gibbon (Sumatra).

Certain generalizations can be made pertaining to all of the Greater Apes: Bodies are short and obese; small eyes, close together, pointing forward; wide and very flaring nostrils with sunken nose-bridge; protuberant muzzle; thin, mobile lips; small, fully opposed thumbs; very long arms; and short hind limbs.

Gorillas are the largest of the Greater Apes and are terrestrial. In walking, they rest partly on the backs of the bent fingers. According to Akeley's observations, they are rather poor climbers. The head of the gorilla is distinguished by the strong jaws and large teeth, the heavy ridges over the eye sockets, and the small ears. Two species are the common gorilla, *Gorilla gorilla*; and the mountain gorilla, *G. beringei.*

Gorillas live in the forests of Africa, with numbers of them found in a wide area of Cameroon and Gabon, north of the Congo, and in Katanga and Zaire. Pockets or "nations" of gorillas also are found in the northwestern portion of Nigeria, and in the volcanic mountains of Kivu. There apparently is no communication between these very isolated communities.

A large male gorilla may stand over five feet high and weigh in excess of 500 pounds (227 kilograms). The lowland dwellers usually have a rusty-gray coat, whereas the mountain dwellers have a black coat. See Fig. 2.

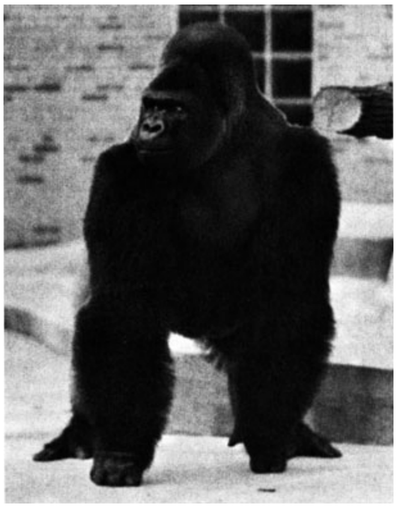


Fig. 2. Lowland gorilla. (*New York Zoological Society.*)

At one time, gorillas would not live long in captivity. This difficulty has been mastered by careful attention to diet and by protecting the animals from the respiratory diseases of humans, to which they are very susceptible. As mentioned by explorer Paul D. du Chaillu, the gorilla has a habit of beating the chest with both hands and displays a very self-centered personality. Psychologist R. M. Yerkes found the gorilla more slow in adaptability to surroundings and also lacking in initiative, when compared with orangutans and chimpanzees of the same age.

An interesting article pertaining to gorilla behavior is "Conversations with a Gorilla," by Francine Patterson (*National Geographic*, **154**, 4. 438–465 (October 1978)). During the past decade, researchers have successfully taught several chimpanzees to converse with signs. In this project, partly funded by the National Geographic Society, Koko is the first gorilla to achieve proficiency. After six years of study, Francine Patterson evaluates Koko's working vocabulary at about 375 signs.

Even in the 1980s, there remains much to be learned concerning the gorilla, particularly pertaining to the behavior of this animal in its natural habitat. There have been numerous observations made, many scientific. some with legendary overtones. Although obviously all communities of gorillas have not been observed let alone located, it has been established that at least some of them construct platform-type homes or nests, built in trees a few feet above ground level. Some authorities have observed true knots used in the doubling over and twisting of tree and root gnarlings in the construction of such platforms. This evidences a superior intelligence which has not been observed in the comparatively few observations of animals in captivity. The strength of the gorilla always has been a topic of some controversy, although all observers agree that it is quite tremendous. There are reports of bending of heavy steel bars and gun barrels. Natives and scientific observers generally agree that gorillas prefer to spend most of their time on the ground, occasionally tree climbing for fruits, and that, when unmolested, are retiring and not ferocious as popularly reported.

Because the chimpanzees most people see on television, in the movies, or at a zoological garden are usually relatively small, sight is lost of the fact that chimpanzees can be quite large, some older male animals rivaling some of the gorillas in proportions. In their natural habitat, they can easily attain a height of five feet and a weight between 150 and 200 pounds (68 and 91 kilograms). Females are usually slightly smaller. These animals live over a large area of Africa, with particular population concentrations around Lake Victoria and in Tanzania. One authority is reported to have observed several hundred of these animals within visual range of a walking path of some 15 miles (24 kilometers) between villages in Cameroon.

In recent years, chimpanzees have become favorites for various types of animal behavior experimentation and observation. E. W. Menzel (*Science*, **182**, 4115, 943–945 (1973)) reports of a study of juvenile chimpanzees, which were carried around an outdoor field and shown up to 18 randomly placed hidden foods. The animals remembered most of the hiding places and the type of food that was in each. Their search pattern approximated an optimum routing, and they rarely rechecked a location that they had already emptied of food. As reported, the chimpanzees appeared to directly perceive the relative position of selected classes of objects and their own position in a scaled frame of reference. They proceeded on the strategy, "Do as well as you can from wherever you are," taking into account the relative preference values and spatial clusterings of the foods as well as distances. As investigator Menzel concludes, "Especially in the light of other recent research, one is struck again by the parallels between chimpanzee and human behavior, the necessity for including representational processes in any adequate formulation of learning and memory, and the apparent evolutionary independence of representational ability and verbal language."

Investigators D. M. Rumbaugh, T. V. Gill, and E. C. von Glasersfeld (*Science*, **182**, 4113, 731–733 (1973)) reported that four studies revealed that a 2½-year-old chimpanzee, after 6 months of computer-controlled language training, proficiently read projected word-characters that constituted the beginnings of sentences and, in accordance with their meanings and serial order, either finished the sentences for reward or rejected them.

The animal is considered to be an extrovert, learning fast from certain types of information and appears to thrive on showing off. However, if it is pushed too hard in learning experiments, tantrums can be expected as well as moodiness. In nature, the male leads the family group as it travels. See Fig. 3.

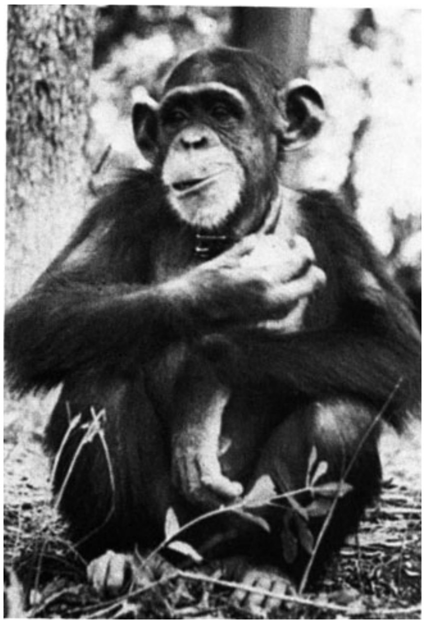


Fig. 3. Chimpanzee. (*A. M. Winchester.*)

In nature or captivity, there appear to be no special breeding seasons. The life span is about 35 years and adulthood is achieved at about 11 years. Chimpanzees feed on fruit, bird eggs, and plant shoots.

In semierect position, the chimpanzee walks on all fours. The arms are used for swinging. In walking, the animal leans on its hands propped by the knuckles, and when it stands erect, its arms reach to just below the knees.

Varied forms of chimpanzees include: The Masked Chimpanzee (*Pan satyrus verus*) found in Upper New Guinea; the Choga (*P. s. satyrus*) found in Lower Guinea; the Koola-Kamba found in Nigeria and environs; and the long-haired Eastern Chimpanzee (*P. s. schweinfurthi*) found in Tanzania and Uganda. But, as mentioned before, these animals are found in many other parts of Africa. In nature, chimpanzees build nests with half-roofs for protection against the elements. These are deftly placed in tree branches several feet above the ground.

Orangutans are found in Borneo and Sumatra. The adult animal may reach a height of 5 to 5½ feet (1.5 to 1.7 meters) and weigh in excess of 150 pounds (68 kilograms). The sharply depressed bridge of the nose accentuates the prominence of the rounded muzzle. For some years, the zoological gardens were more successful in boarding orangutans than gorillas. However, the orangutans can be considered delicate when in captivity, particularly in northern climates. These animals, like the chimpanzee to some extent, have been found to be very intelligent, but they are less desirable as subjects because of their uncertain temper, a characteristic particularly true of older males. The term orangutan (or *Orang-Utan*) is Malayan for "man of the woods."

Orangutans are usually brownish-red in coloration, with a broad, flat face, and long, coarse hair. A sac extending from the front of the throat extends to the armpits. This feature, plus the depressed nose-bridge, the huge cheek flaps, comparatively small legs, disproportionately long and heavier forelimbs, and obese stomachs, particularly in older specimens, gives them a rather grotesque appearance. See Fig. 4. If adopted into a human household, the orangutan demonstrates amazing adaptability often showing, as with a dog for example, definite preferences for certain people, while also demonstrating a marked antipathy for others.

In nature, the animals are vegetarians, preferring fruit (particularly of the palm nut tree), but tolerate a wider diet when in captivity. They also are "sleeping platform" builders as some of the other apes described. Even the heavier adults can travel rapidly among the tree tops and display brachiation as previously defined. See also **Mammalia**; and **Primates.**

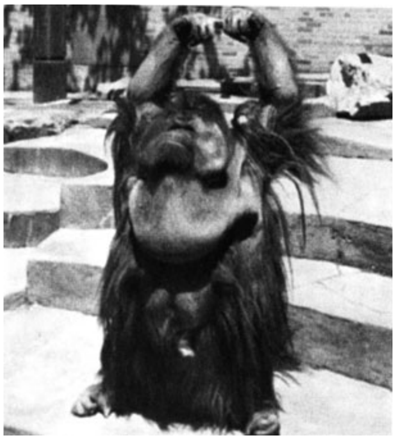


Fig. 4. Adult orangutan. (*New York Zoological Society.*)

Additional Reading

Biewener, A. A.: "Biomechanics of Mammalian Terrestrial Locomotion," *Science*, 1097 (November 23, 1990).

Booth, W.: "Chimps and Research: Endangered?" *Science*, 777 (August 12, 1988).

Bower, B.: "Biographies (Chimpanzee) Etched in Bone," *Science News*, 106 (August 18, 1990).

Galdikas, B. M. F.: "Orangutan Tool Use," *Science*, 152 (January 13, 1989).

Gibbons, A.: "Chimps: More Diverse Than a Barrel of Monkeys," *Science*, 287 (January 17, 1992).

King, F. A.: "Primates," *Science*, 1475 (June 10, 1988).

Morell, V.: "A Hand on the Bird — And One on the Bush (A Controversial New Theory Holds that Nonhuman Primates are 'Handed' Just as Humans are)," *Science*, 33 (October 4, 1991).

Ristau, C. A., Editor: "Cognitive Ethology: The Minds of Other Animals," Erlbaum, Hillsdale, New Jersey, 1991.

Vessels, J.: "Koko's Kitten," *Nat'l. Geographic*, 110 (January 1985).

ANTIBIOTIC. A biochemical drug, derived from one or more kinds of microorganisms, which has the ability to (1) inhibit the growth (*bacteriostatic agent*), or (2) to kill (*bactericidal agent*) a number of other microorganisms and thus of immense value in treating a number of diseases that result from microbial infection. Major antibiotics administered today are listed in Table 1.

TABLE 1. CLASSIFICATION OF DRUGS USED IN THE CHEMOTHERAPY OF MICROBIAL DISEASES

Bacteriostatic Agents	Bactericidal Agents
Sulfonamides	Penicillins (penicillin G, penicillin V, methicillin, oxacillin, cloxacillin, nafcillin, ampicillin, amoxicillin, carbenicillin)
Trimethoprim	Cephalosporins
Tetracyclines	Aminoglycosides (streptomycin, neomycin, kanamycin, gentamicin, tobramycin, amikacin)
Chloramphenicol	Vancomycin
Erythromycin	Polymyxins (polymyxin B, colistin)
Lincomycin or clindamycin	Bacitracin

Effective chemotherapy in the use of antibiotics depends upon *selective toxicity*. This may be defined as the ability of the drug to inhibit microorganisms at concentrations tolerated by the host. It has been found that those antimicrobial agents which are most effective target their action on the anatomic structures or biosynthetic functions that are unique to microorganisms. Some antibiotics interfere with the synthesis of the mucopeptide layer of the bacteria cell wall. This structure is not present in the cells of the host. The penicillins, cephalosporins, and vancomycin act in this manner. Other drugs, such as colistin and polymyxins, alter the permeability of the bacterial cell membrane. This allows the cell contents to leak out. The aminoglycoside antibiotics and the tetracyclines are believed to act upon a ribosomal subunit, thus interfering with the target microorganism's ability to synthesize protein at the translational level. Chloramphenicol, erythromycin, and lincomycin act in a similar manner, but on a different subunit of the bacterial ribosome. Rifampin selectively inhibits bacterial DNA-dependent RNA polymerase, but fortunately does not affect this same enzyme in the host cell. Inhibition of the microbial synthesis of folic acid and of the precursors of folic acid are the principal actions of the sulfonamides.

With a wide variety of antimicrobial drugs available, the following guidelines are frequently followed in determining the most effective drug to use: (1) The infecting organism must be identified; (2) consider the antimicrobial susceptibility pattern of the microorganism; (3) consider the realtive merits of bactericidal versus bacteriostatic drugs; (4) define the site of infection; (5) consider the possible advantages of using a combination of antibiotics; (6) consider the clinical pharmacology, including dosages, route of administration, possible adverse reactions, drug interactions, serum levels, among other factors; (7) estimate the optimal duration of therapy; and (8) if the drug(s) are to be administered in a hospital, check if there are any limitations on the use, particularly of new drugs.

Development of Antimicrobial Drugs

The first scientific demonstration of microbial antagonism was made by Pasteur and Joubert in 1877 when they observed that certain common bacteria inhibited the growth of anthrax bacilli. This basic phenomenon by which one microorganism destroys another to preserve its own life was, at that time, called *antibiosis* by Vuillemin (1889). In the decades that followed, the therapeutic efficacy of antibiotics to control infectious disease was eventually demonstrated, after which the pursuit of microbial antagonists rapidly became an organized applied science.

Pyocyanase was the first microbially derived antibiotic product to be used in treating bacterial infections in humans. Although it had only limited clinical use, it is interesting historically because it demonstrated as early as 1906 the principle of selective toxicity, i.e., specificity of action against the invading pathogen and a correlative lack of toxic action in the host. Following the decline in use of pyocyanase, it was almost a quarter of a century before interest in anti-infective agents from microbial sources was renewed.

In 1929, the British bacteriologist Alexander Fleming published his observations on the inhibition of a staphylococcus culture by growing colonies of *Penicillium notatum*. This report went largely unpursued for a decade, after which Florey and Chain reinvestigated Fleming's work and, in 1941, demonstrated the clinical usefulness of penicillin. In 1939, Dubos, by careful, well-planned studies, obtained the antibiotic tyrothricin from the soil organism *Bacillus brevis*. Although tyrothricin found only limited use, the work of Dubos on the chemical, biological, and physical properties of this antibiotic contributed immensely toward forcing a realization of the potentialities of antibiotic substances. Similarly, Waksman undertook a systematic search for antimicrobial substances in a group of soil-inhabiting microbes known as *Streptomyces* and announced the discovery of streptomycin in 1944.

The foregoing discoveries stimulated worldwide interest, and the discovery of useful new antibiotics during the period 1939–1959 was prolific. During this period, the major classes of antibacterial antibiotics were recognized. Many specific drugs which presently occupy places in therapeutic practice were discovered directly in microbial fermentations. The later work in the field had consisted principally of chemical modifications of antibiotic substances previously known.

See separate article on **Sulfonamide Drugs**.

Classification of Antibiotics

The antibiotics comprise a widely diverse group of substances, differing not only in chemical structure, but also in their mode of action, antibacterial spectra, origin, and other features. Grouping of antibiotics in accordance with their toxic action on target microorganisms, as previously mentioned, bears little resemblance to their classification according to origin and chemical structure. The latter classification is presented in Table 2. Among the most important of the antibiotics used in current medical practice are: (1) the beta-lactams, which include the penicillins and cephalosporins; (2) the tetracyclines; (3) the macrolides; and (4) the aminoglycosides.

TABLE 2. CLASSES OF ANTIBIOTICS ON BASIS OF BIOGENETIC ORIGIN AND CHEMICAL STRUCTURE

Amino Acid Units	Acetate/Propionate	Sugar Units
Amino acid cogeners	Fused-ring systems	Aminoglycosides
D-Cycloserine	Tetracyclines	Streptomycin
Chloramphenicol	Oxytetracycline	Kanamycins
Beta-Lactams	Chlortetracycline	Gentamicins
Penicillins	Steroidal antibiotics	Neomycins
Cephalosporins	Fusidic acid	Tobramycins
Polypeptides	Griseofulvin	Amikacins
Bacitracins	Antibacterial macrolides	
Polymixins	Erythromycin	
Viomycin	Oleandomycin	
Capreomycin	Leucomycins	
Vancomycin	Spiramycins	
	Polyene macrolides	
	Nystatin	
	Amphotericins	
	Ansa-macrolides	
	Rifamycins	

Note: Some of the foregoing are mainly of historical or research interest.

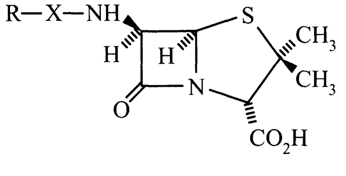
Penicillins. The penicillins are chemically characterized by a four-membered lactam ring fused to a thiazolidine ring and are differentiated by the side-chain (R) attached to the bicyclic nucleus. See Table 3. Penicillins are sometimes named by attaching the chemical name of the R-substituent as a prefix to the word penicillin. Thus, in the case where R is $C_6H_5CH_2$—, the compound may be called *benzylpenicillin*. However, this compound in commerce is more commonly referred to as *Penicillin G*. Similarly, in the case where R is $C_6H_5OCH_2$—, the compound may be called *phenoxymethylpenicillin*, although commercially it is more commonly called *Penicillin V*. Most frequently, the commercial names bear no resemblance to structure.

The naturally occurring penicillins have been discovered in the fermentation broths of *Penicillium* and *Cephalosporium* cultures.

The earliest of the penicillins, simply called penicillin, was benzylpenicillin (Penicillin G). This early product, still used, was shown to have several limitations, including acid instability, allergenicity, and susceptibility to enzymatic inactivation by penicillinases.

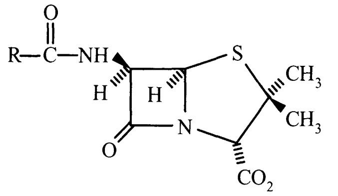
In 1947, it was discovered that addition of phenylacetic acid to penicillin fermentation media increased the yield of benzylpenicillin at the expense of other less desirable natural penicillins. Following this observation, a new generation of *biosynthetic* penicillins was prepared by addition of monosubstituted acetic acid derivatives to penicillin fermentations. The most important of these biosynthetic derivatives is penicillin V, obtained by adding phenoxyacetic acid to penicillin growth media. This widely used antibiotic is relatively stable in dilute acid, is not destroyed by the acidic contents of the stomach, and consequently can be effective by oral administration.

Although the biosynthetic approach created several new penicillins, it was limited in the type of side-chain (R) that could be introduced. Only derivatives with an unsubstituted methylene adjacent to the amide carbonyl (X as shown below) could be generated.

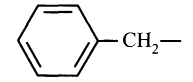
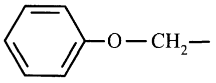
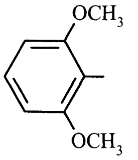
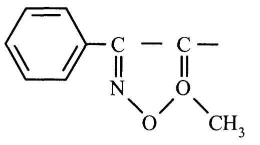
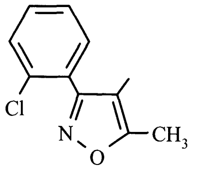
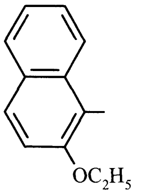
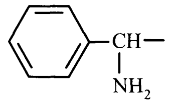
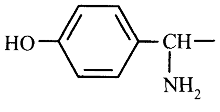
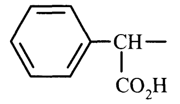


The next major breakthrough in penicillin research came in 1959 with the isolation of the penicillin nucleus 6-aminopenicillanic acid (6-

TABLE 3. STRUCTURES OF SOME OF THE PRINCIPAL PENICILLINS



Portion of β-lactam ring cleaved by penicillinase (β-lactamase)

Generic or Chemical Name	Substituent Side Chain (R)
Penicillin G (Benzylpenicillin)	
Penicillin V (Phenoxymethylpenicillin)	
Methicillin	
Oxacillin	
Cloxacillin	
Nafcillin	
Ampicillin	
Amoxicillin	
Carbenicillin	

APA) from fermentation mixtures to which no side-chain precursor had been added. Although chemical synthesis of 6-APA and its utility for the preparation of new penicillins by acylation were announced by Sheehan in 1958, the fermentation method provided the first practical means of obtaining large quantities of 6-APA. Chemical acylation of 6-APA allowed introduction of almost unlimited varieties of side chains and gave rise to a third generation of penicillins called *semisynthetic penicillins.*

The nature of the acyl side chain has been found to have a profound effect on the properties of the penicillins, influencing such therapeutically important properties as acid stability, oral absorption, serum protein binding, penicillinase resistance, and gram-negative activity.

One of the major developments resulting from the availability of 6-APA was the creation of semisynthetic penicillins that resist destruction by the penicillinases. The empirical finding that triphenylmethylpenicillin was resistant to penicillinase led to the screening of other penicillins with sterically hindered side-chains, partly because the presence of a bulky group near the beta-lactam ring resulted in reduced affinity of these substances for the enzyme. Methicillin and cloxacillin are compounds with such side-chains which have proved clinically useful.

Ampicillin. This drug has a broader range of activity than that of penicillin G. The spectrum encompasses not only pneumococci, meningococci, and a number of streptococci, but also several gram-negative bacilli.

Amoxicillin. As evident from Table 3, amoxicillin is structurally similar to ampicillin with exception of an OH instead of an H in one of the positions of the side-chain. Although this difference does not alter the spectrum of the two drugs, amoxicillin is better absorbed from the gastrointestinal tract, resulting in longer effective concentrations of the drug present in the circulation.

Carbenicillin. This drug has a carboxyl rather than an amino substituent and has greater activity against gram-negative bacilli.

Methicillin. This drug is administered intravenously or intramuscularly. In recent years, the semisynthetic penicillinase-resistant oxacillin and nafcillin have markedly supplanted methicillin for many situations.

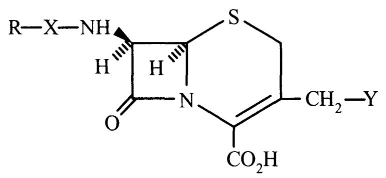
Cephalosporins. These drugs constitute another major class of beta-lactam antibiotics and are chemically characterized by a beta-lactam fused to a dihydrothiazine ring. In contrast to the penicillins, where the side-chain of the antibiotic varies, depending upon precursors present in the fermentation mixture, fermentation-derived cephalosporins contain the same side-chain. *Cephalosporin C*, the parent antibiotic of this class, is not useful clinically. For example, it is about 0.1% as active as benzylpenicillin against staphylococci. However, in early research, cephalosporin C exhibited certain interesting properties which provoked further study. Cephalosporin C was more stable toward acid than penicillin; it was unaffected by penicillinases; it exhibited appreciable activity against some gram-negative bacteria, and it appeared to have no cross-allergenicity with the penicillins. Consequently, in the late 1950s, many laboratories investigated both chemical and microbiological methods for removing the aminoadipoyl side-chain of cephalosporin C to obtain the cephalosporin nucleus, 7-aminocephalosporanic acid (7-ACA).

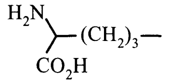
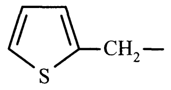
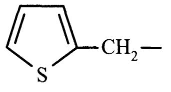
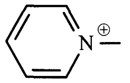
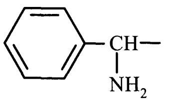
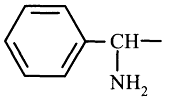
A practical chemical process for accomplishing this transformation was announced in 1962. Like 6-APA, previously described, 7-ACA can be readily acylated, and a large number of semisynthetic cephalosporins thus have been possible. Cephalothin was the first clinically useful broad-spectrum cephalosporin to emerge from synthetic studies. Cephaloridine soon followed cephalothin and was found to be 2 to 8 times more active than the latter against gram-positive organisms. In 1970, cephaloglycin became commercially available as the first orally effective broad-spectrum cephalosporin. Cephalexin is metabolically more stable than cephaloglycin. More recently available cephalosporins have included cephapirin, cephradine, and cefazolin.

The basic structure of the cephalosporins and structures of several of the antibiotics in this family are shown in Table 4.

Aminoglycosides. These drugs comprise a class of potent broad-spectrum antibiotics which are chemically characterized by basic carbohydrate moieties glycosidically bound to a cyclitol unit. In general, the aminoglycosides are effective against most gram-positive and gram-negative bacteria, as well as *Mycobacterium tuberculosis*. Because of their highly ionic nature, the aminoglycosides are not absorbed

TABLE 4. STRUCTURES OF REPRESENTATIVE CEPHALOSPORINS

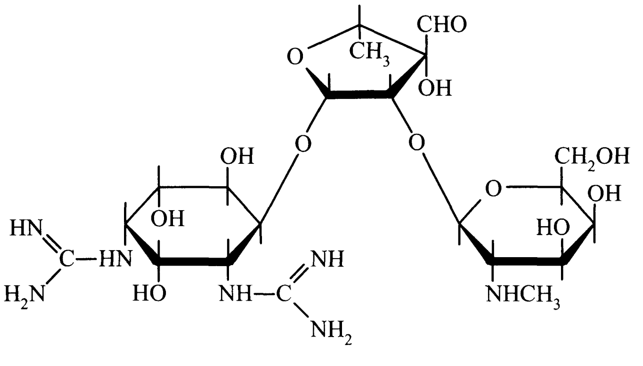


Generic or Chemical Name	R	X	Y
Cephalosporin C (parent of class—essentially inactive)		C=O	CH_3CO_2—
7-Aminocephalosporanic acid	—	H	CH_3CO_2—
Cephalothin		C=O	CH_3CO_2—
Cephaloridine		C=O	
Cephaloglycin		C=O	CH_3CO_2—
Cephalexin		C=O	H

Note: Important cephalosporins not included in table include Cefamandole, cefazolin, Cefoxitin, cephapirin, cephradine.

from the gastrointestinal tract and must be administered parenterally. In a small percentage of patients, prolonged use of this class of antibiotics can adversely affect the eighth cranial nerve, causing some impairment of hearing and balance.

Streptomycin. Discovery in 1944 of streptomycin (structure shown below) drew immediate interest because it was the least toxic of the broad-spectrum antibiotics known at that time. Indeed, streptomycin was used to treat many gram-negative microbial infections, but because of the ease with which organisms developed resistance to it during treatment, many of these applications were abandoned when the tetracyclines, discussed later, became available. Streptomycin was the first parenterally administered antibiotic active against many microorganisms, but during the last several years, its use is limited essentially to three situations: (1) the initial treatment of serious tuberculous infections when the principal drugs of choice (isoniazid, rifampin) cannot be used because of their adverse effects on a particular patient; (2) treatment of enterococcal and other infections in which synergism between a penicillin and an aminoglycoside is desired; and (3) treatment of certain uncommon infections (plague and tularemia).



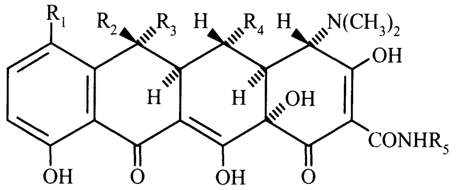
Kanamycin. Considerably broader in spectrum than streptomycin, kanamycin is more effective against gram-negative bacilli (other than *Pseudomonas*) and also is effective to a degree against *Staph. aureus*. However, it is ineffective against streptococci and pneumococci. The availability of penicillinase-resistant penicillins and cephalosporins essentially obsoleted kanamycin as the primary drug in the treatment of staphylococcal infections. Kanamycin has been essentially replaced by gentamicin and other aminoglycosides which are less ototoxic (adverse to hearing), and which also have a wider range of antibacterial activity.

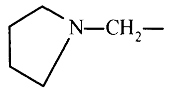
Gentamicin. One of the successors of kanamycin, gentamicin possesses essentially the same spectrum as kanamycin, but is also active against *Pseudomonas aeruginosa*. An advantage of gentamicin is its penetration into pleural, ascitic, and synovial fluids where there is inflammation. Although not necessarily the drug of choice, gentamicin has been used in the treatment of acute cholecystitis, acute septic arthritis, anaerobic infections, *Bacillus* infections, gram-negative bacteremia, infective endocarditis, meningitis, osteomyelitis, peritonitis, staphylococcal infections, and tularemia, among others. In some situations, gentamicin acts synergistically with penicillin.

Tobramycin. Pharmacologically, tobramycin is quite similar to gentamicin. The drug is somewhat more active against *Ps. aeruginosa* than gentamicin. Tobramycin also acts synergistically with penicillin, but to a lesser degree than gentamicin.

Amikacin. This drug is a semisynthetic derivative of kanamycin. It is much less sensitive to the enzymes that inactivate aminoglycoside antibiotics. The spectrum is similar to that of gentamicin. Amikacin principally finds use in the treatment of infections arising from bacteria that are resistant to gentamicin and/or tobramycin.

TABLE 5. STRUCTURES OF REPRESENTATIVE TETRACYCLINES



Generic or Chemical Name	R_1	R_2	R_3	R_4	R_5
Tetracycline	H	OH	CH_3	H	H
Chlortetracycline	Cl	OH	CH_3	H	H
Oxytetracycline	H	OH	CH_3	OH	H
Demethylchlortetracycline	Cl	OH	H	H	H
Methacycline	H	$=CH_2$	$=CH_2$	OH	H
Doxycycline	H	H	CH_3	OH	H
Rolitetracycline	H	OH	CH_3	H	

Note: Minocycline is not included in table.

Tetracyclines. These drugs comprise a family of broad-spectrum antibiotics possessing a common perhydronaphthacene skeleton. They have a wider range of antimicrobial activity than other classes of clinically useful antibiotics. See Table 5. The tetracyclines are active against many species of gram-positive and gram-negative bacteria, spirochetes, rickettsiae, and some of the larger viruses.

Chlortetracycline. The first member of this class to be isolated, chlortetracycline, was discovered in 1948 among the metabolites of *Streptomyces aureofaciens.* Oxytetracycline was isolated two years later from a *S. rimosus* fermentation. Both antibiotics quickly found wide medical use, not only because they were effective orally, but because they were useful against a much wider spectrum of bacteria than penicillin G.

Chemical studies on chlortetracycline and oxytetracycline, which provided a basis for structure assignment, in general led to products with diminished or no antibacterial activity. In 1953, the first scientific reports appeared describing an active tetracycline prepared by chemical modification of a fermentation product. This was *tetracycline*, the parent member of this family of antibiotics, prepared by catalytic hydrogenolysis of chlortetracycline. It was more stable and better tolerated than its fermentation-produced progenitor, and almost completely displaced chlortetracycline from medical practice. Interestingly, tetracycline was later found in fermentation broths of a mutant strain of *S. aureofaciens* and also may be manufactured by this method.

Following the discovery of tetracycline, useful new drugs from chemical modification of tetracycline antibiotics were slow in coming, for the complexity and chemical lability of the tetracyclines did not render them amenable to facile systematic studies of the relationships between chemical structure and biological properties. Unlike the β-lactam antibiotics, where structural modifications were being sought mainly to improve their antibacterial spectra and potency, superior semisynthetic tetracyclines were obtained, in general, with improved pharmacokinetic properties, i.e., through such factors as rate of oral absorption, degree of serum protein binding, rate of urinary excretion, and biological half-life.

An effective approach to the discovery of superior tetracycline antibiotics stemmed from studies yielding tetracyclines modified at the C-6 position. Thus, demethylchlortetracycline and methacycline were found to be somewhat superior to tetracycline in terms of a longer serum half-life, and later compounds, such as doxycycline, were shown to exhibit near-ideal pharmacokinetics. Doxycycline, among a number of other tetracyclines, remains in wide use today.

Among the many diseases treated with tetracyclines are urinary tract infections, gonorrhea, nongonococcal urethritis, Rocky Mountain spotted fever, other rickettsioses, mycoplasmal pneumonia, chlamydial diseases (psittacosis, trachoma, lymphogranuloma venereum), brucellosis, plague, cholera, granuloma inguinale, syphilis, and gonococcal pelvic inflammatory diseases, particularly in a number of instances where a patient may be allergic to penicillin. Tetracyclines also have been used in the treatment of cystic acne.

Doxycycline has been reported as effective in the prophylaxis of traveler's diarrhea in Kenya. The tetracycline minocycline is sometimes used instead of the sulfonamides against noncardial infections. It is not used in connection with meningococcal infections.

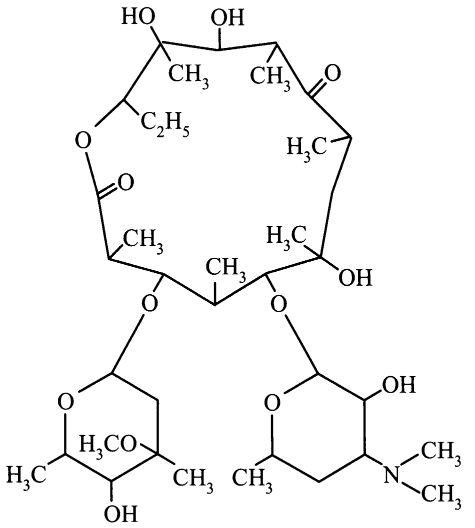
Macrolides. The macrolides comprise a family of antibiotics chemically characterized by a macrocyclic lactone to which one or more sugars are attached. The compounds are often divided into various subgroupings, but the group with antibacterial properties are known as *antibacterial macrolides.* They are distinguished chemically by having, in addition to the large lactone, various ketonic and hydroxyl functions and glycosidically bound deoxy sugars. A second grouping of commercial importance is known as the *polyene macrolides*, chemically characterized by extended conjugated double-bond systems. The polyenes are devoid of antibacterial activity, but are potent antifungal agents.

A number of the antibacterial macrolides have been found to be clinically useful chemotherapeutic substances, falling generally under the title of *medium-spectrum antibiotics.* This term is taken to mean that these substances are effective against most gram-positive bacteria and have a degree of activity against certain gram-negative organisms, such as *Haemophilus*, *Brucella*, and *Neisseria* species. The antibacterial macrolides also appear to inhibit certain pleuropneumonialike organisms.

Erythromycin. This is the principal drug in this category. Although available as the parent entity, semisynthetic derivatives have proved to be clinically superior to the natural cogener. Like the tetracyclines, synthetic transformations in the macrolide series have not significantly altered their antibacterial spectra, but have improved the pharmacodynamic properties. For example, the propionate ester of erythromycin lauryl sulfate (erythromycin estolate) has shown greater acid stability than the unesterified parent substance. Although the estolate appears in the blood somewhat more slowly, the peak serum levels reached are higher and persist longer than other forms of the drug. However, cholestatic hepatitis may occasionally follow administration of the estolate and, for that reason, the stearate is often preferred.

Erythromycin is effective against Group A and other nonenterococcal streptococci, *Corynebacterium diphtheriae*, *Legionella pneumophila*, *Chlamydia trachomatis*, *Mycoplasma pneumoniae*, and *Flavobacterium.* Because of the extensive use of erythromycin in hospitals, a number of *Staph. aureus* strains have become highly resistant to the drug. For this reason, erythromycin has been used in combination with chloramphenicol. This combination is also used in the treatment of severe sepsis when etiology is unknown and patient is allergic to penicillin.

A structural representation of erythromycin is shown as follows.



Other Antibiotics

Chloramphenicol. This compound is derived from *Streptomyces venezuelae* or by organic synthesis. It was the first substance of natural origin shown to contain an aromatic nitro group. Although the drug is a valuable broad-spectrum antibiotic, its use has been somewhat limited because of the occasional development of aplastic anemia in the patient. Thus, its use has been largely confined to its administration as the drug of choice in patients allergic to penicillin in connection with typhoid fever, nontyphoidal salmonelloses (due to ampicillin-resistant strains), *H. influenzae* meningitis, meningitis arising from *N. meningitidis*, and *Str. pneumoniae*. Because effective and safer drugs are not available, it is used for infections arising from *Bact. fragilis*. The drug is administered orally or intravenously.

Vancomycin. This is a narrow-spectrum antibiotic and produced by *Streptomyces orientalis* or synthetically. Its effectiveness is essentially confined to the treatment of streptococci (including enterococci), pneumococci, staphylococci, and a few other gram-positive bacteria. Serious side effects include possible hearing loss and renal insufficiency, particularly when the drug is administered with an aminoglycoside.

Polymyxins. This is a generic term for a series of antibiotic substances produced by strains of *Bacillus polymyxa*. Various polymyxins are differentiated by letters A, B, C, D, and E. All are active against certain gram-negative bacteria. Polymyxin B and E (colistin) have been the most important in the past, but currently are only rarely used—because they have been replaced by more effective aminoglycosides. The B and E drugs are effective against most of the common aerobic gram-negative bacilli, but not *Proteus*, *Providencia*, and *Serratia*. Prior to their replacement by aminoglycosides, the polymyxins were used mainly in connection with infections arising from *Ps. aeruginosa*.

Spectinomycin. This drug finds principal application in the treatment of gonorrhea. It should be noted that the antibiotic resistance among *N. gonorrhoeae* has caused a number of therapeutic problems. It has been found that only by escalating the antibiotic doses and using probenecid to retard the excretion of penicillin and ampicillin (the drugs of choice) has the continued effective use of penicillin, ampicillin, and tetracycline been possible. Even with modifications in the therapy, from 3 to 8% of cases fail to respond to the usual regimens for uncomplicated gonorrhea. Thus, the treatment of uncomplicated gonorrhea that fails to respond to the usual regimen is spectinomycin therapy.

Chemoprophylaxis with Antibiotics

In addition to their use in treating infections arising from bacteria and other microorganisms, antibiotics are also used in chemoprophylaxis, i.e., treating a patient before or shortly after the entry of pathogenic organisms. There are three common situations: (1) preventing infection following exposure to known pathogens; (2) preventing specific infections in highly susceptible individuals; and (3) preventing postoperative infectious complications.

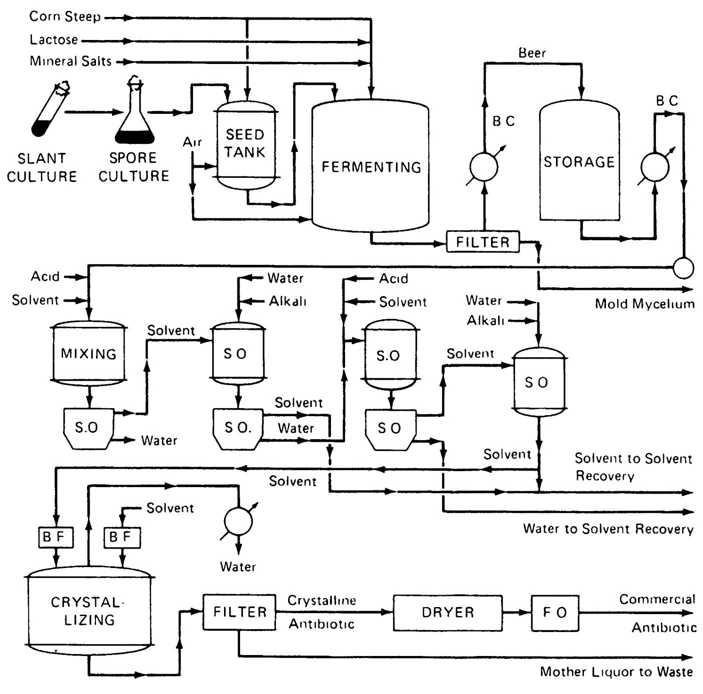
Antibiotics in Feedstuffs

For several years, antibiotics have been used in feedstuffs, not only to lower the incidence of certain diseases in livestock, but antibiotics also play a function in the rate of growth of animals. As of the late 1980s, the question of whether or not the use of antibiotics in livestock may have an adverse effect on human consumers of meat remains unresolved.

Manufacture of Antibiotics

Generally, most antibiotics and the starting materials for semisynthetic antibiotics are manufactured by fermentation, with accompanying extraction, purification, crystallization, and packaging operations. Commercial fermenting vessels are stainless- or carbon-steel enclosed tanks with capacities up to several tens of thousands of gallons (many hundreds of hectoliters). Such factors as aeration, agitation, temperature, and pH must be monitored and controlled carefully.

The antibiotic-producing microorganism is grown in submerged culture in a fermentation medium which contains various carbon, nitrogen, and trace-metal sources, required by the organism for its nutrition. The organism is grown under conditions of pure culture, that is, other microorganisms excluded from the fermentation inasmuch as the latter will compete for nutrients, may contribute undesirable contamination and reduce yield of the desired product. When the fermentation has reached peak potency (varies with each product), the antibiotic may be recovered by an extraction technique, such as distribution into a water-immiscible solvent, ion-exchange chromatography, or precipitation. Following extraction, purification and crystallization are carried out by procedures compatible with the physicochemical properties of the particular antibiotic being produced. A representative flowsheet is shown in the accompanying figure.



Representative schematic of materials flow in commercial antibiotic manufacturing process. B.C. = brine cooler; S. O. = separating operation; B.F. = bacteriological filter.

Additional Reading

Gilpin, R. K., and L. A. Pachla: "Analysis for Antibiotics," *Analytical Chemistry*, 130R (June 15, 1991).

Jacoby, G. A., and G. L. Archer: "New Mechanisms of Bacterial Resistance to Antimicrobial Agents," *N. Eng. J. Med.*, 601 (February 28, 1991).

Mark, A. L.: "Cyclosporine, Sympathetic Activity, and Hypertension," *N. Eng. J. Med.*, 746 (September 13, 1990).
Moberg, C. L., and Z. A. Cohn, Eds.: "Launching the Antibiotic Era," Rockefeller University Press, New York, 1990.
Moberg, C. L., and Z. A. Cohn: "Rene Jules Dubos," *Sci. Amer.*, 66 (May 1991).
Moberg, C. L.: "Penicillin's Forgotten Man: Norman Heatley," *Science*, 734 (August 16, 1991).
Neu, H. C.: "The Crisis in Antibiotic Resistance," *Science*, 1064 (August 21, 1992).
Reese, K.: "The Road to Peoria and Penicillin Production," *Today's Chemist at Work*, 48 (August 1992).
Swan, H. T.: "The Antibiotic Record," *Science*, 1387 (March 23, 1990).

ANTIBODY. This article gives a generalized description of antibodies and their role in the body's immune system. More details in terms of the most recent findings in this field are given in the entry on **Immune System and Immunology.** In medicine and physiology, immunity is the ability of the body to resist invasion by pathogenic organisms and substances. Immunity may be initially in place, that is, genetically ordered for a given species. Humans are naturally immune to canine distemper; dogs are immune to measles; rats are immune to diphtheria; and domestic fowls are immune to anthrax. Many other examples could be cited. Immunity may be acquired as the result of exposure to an invasive pathogen, triggering the immune system to construct cells that will be in reserve to resist subsequent invasions by the same pathogen. Immunity also may be acquired artifically through the use of preventive immunization techniques. Immunity is effected through antibodies.

Any substance that can provoke a response by the body's immune system is called an *antigen*. This property of an antigen is referred to as *immunogenicity*. Although the first antigens to be investigated were microorganisms and proteins foreign to the body, research during recent decades has been directed toward understanding the immune response at the molecular level—for it is at this level that the actions and reactions of the immune system occur. At the molecular level, numerous previously unsuspected complexities of the immune process have been revealed and still others are only partially understood, if at all. It has been discovered that several cell types, in addition to the lymphocytes, act cooperatively in effecting what might be called the total immune response. Although the lymphocytes appear to play the dominant role in the immune system, several other cells are now known to cooperate with the lymphocytes, and the functions of these other cells are no longer considered of secondary importance. Study of the antigens at the molecular level also has contributed to a much better understanding of the immune response.

Numerous molecules can evoke an immune response, sometimes when the responses from the standpoint of protecting body functions are not immediately obvious, other than that such molecules do not meet the criteria of "self" and "nonself" described later. In a general way, it may be observed that the immune system tends to have a bias toward suspicion and may, on occasion, overreact, as in cases of autoimmunity. Currently, it is generally hypothesized that recognition of antigens at the receptor sites of the lymphocytes is based upon the shapes of molecules, reminiscent of some of the current hypotheses concerning taste and odor receptors in the tongue and nasal membranes. Until the mechanism occurring at the receptor sites is more fully explained, numerous questions as regards what molecules do and do not evoke immune response will remain unanswered.

Considerable research with synthetic polymers comprised of various amino acids has been undertaken in an effort to determine the requirements of immunogenicity. It has been established, for example, that tyrosine as well as some other aromatic amino acids will confer immunogenicity to certain polypeptides which in themselves are not or are only slightly antigenic. Further, it has been found in such cases that the antibody is directed against the polypeptide and not the amino acid. Research with synthetic polymers led investigators to the finding that immune response is under genetic control, this based upon the observations that different animal strains and species respond differently to a given polymer. However, in a descriptive fashion, this principle had been demonstrated by the different reactions to antigens by various species many decades ago.

Although not catalytic in the usual sense, certain substances, known as *adjuvants*, are capable of enhancing the immunogenicity of certain antigens. Among the adjuvants are aluminum salts, bacterial endotoxins, bacillus Calmette-Guérin (BCG), *Bordetella pertussis*, and mycobacteria. These materials and this phenomenon have been important in immunity research. Sometimes adjuvants are used clinically in connection with certain immunizations, such as against tetanus.

Where antigens are introduced into the body intravenously, they usually travel rapidly to the spleen, followed by the fast production of an antibody. Subcutaneous or intradermal injection of antigens most frequently localize in the lymph nodes and antigens that are inhaled favor local sensitization. In some cases, such as tetanus immunization, toxin produced by the bacteria may be slow and insufficient to provoke a significant immunologic reaction. Thus, the requirement for properly timed booster injections.

Clinical Use of Antigens. Without the benefit of understanding the complexities of the immune system, particularly at the molecular level, much progress was made over the years in taking advantage of certain antigens and a qualitative or descriptive understanding of the immune response. Thus, the early development of vaccines and antitoxins.

The history of the development of antitoxins in combating bacterial infection dates back to the early beginnings of organized bacteriology. Behring was the first to show that animals that were immune to diphtheria contained, in their serum, factors which were capable of neutralizing the poisonous effects of the toxins derived from the diphtheria bacillus. While this work was carried out in 1890, prior to many of the great discoveries of mass immunization, and much later the antibiotics, it is interesting to note that there remains a place for antitoxins, even though relatively limited in modern medical treatment or prophylaxis of a few diseases, such as tetanus, botulism, and diphtheria. In the case of diphtheria, equine antitoxin is the only specific treatment available. However, it is only reasonably effective if used during the first 48 hours of the onset of the disease. Trivalent (ABE) antitoxin is used in the treatment of botulism. In the treatment of tetanus, human tetanus immune globulin is preferred, but when it is not available, equine antitoxin is substituted. See **Antitoxin**.

For many years the preferred approach to immunity to infectious disease has been by development of active immunity through the injection of a vaccine. The vaccine may be either an attenuated live infectious agent, or an inactivated or killed product. In either case, protective substances called antibodies are generated in the bloodstream; these are described in the next section. Vaccines for a number of diseases have been available for many years and have assisted in the erradication of some diseases, such as smallpox. As new strains of bacteria and viruses are discovered, additional vaccines becomes available from time to time. See **Vaccine.**

Antibodies. Antigens are excluded from the body by skin and mucous membranes. If these barriers are penetrated, the foreign organism may be ingested by phagocytic cells (monocytes, polymorphs, macrophages) and subsequently destroyed by cytoplasmic enzymes. Some time after a foreign macromolecule has entered the body, induced mechanisms come into play. There are two basic biological manifestations of the immune reaction: (1) Immunity to infectious agents; and (2) specific hypersensitivity. Hypersensitivity, or the heightened response to an agent, can be divided into anaphylactic, allergic, and bacterial. Anaphylaxis, which can be produced by either active or passive sensitization, is a laboratory tool for studying the fundamental nature of hypersensitivity. The amounts of antigen and antibody involved, as well as the nature and source of the antibody, govern the extent of the reaction.

In immunity to infectious agents, some time after a foreign macromolecule has entered the body, induced mechanisms come into play, which result in the synthesis of specially adapted molecules (*antibodies*) capable of combining with the foreign substances which have elicited them. Most macromolecules (proteins, carbohydrates, nucleic acids, etc.) can function as antigens, provided that they are different in structure from autologous macromolecules, i.e., from the macromolecules of the responding organism.

Antibodies are proteins with a molecular weight of 150,000–1,000,000 and with electrophoretic mobility predominantly of gamma globulins. The combination between antigen and antibody results in inhibition of the biological activity of the antigen and leads to increased rate of ingestion (*opsonization*) of the antigen by phagocytic cells. In addition, combination of antigen and antibody results in the activation of a complex chain of interacting constitutive molecules—the *complement system*—leading to lysis of the cell membranes to which antibody, directed against cellular antigens, is attached.

Biochemical Individuality. This is a unique quality, genetically determined, for each individual and is exhibited with respect to: (1) The composition of blood, tissues, urine, digestive juices, cerebrospinal fluid, etc.; (2) the enzyme levels in tissues and in body fluids, particularly the blood; (3) the pharmacological responses to specific drugs and poisons; (4) the biochemical responses to bacteria, fungi, and other microorganisms; (5) the quantitative needs for specific nutrients—minerals, amino acids, vitamins, etc.—and in a number of other ways, including reactions of taste and smell and the effects of heat, cold, and electricity. Although individual *similarities* are readily apparent at the macro level, each individual must possess a highly distinctive pattern, since the differences between individuals with respect to measurable items in a potentially long list are by no means trifling. Out of these relatively small, but distinct differences has risen the concept of the so-called normal individual, against which high and low levels of response are compared when considering the "chemistry" of a given individual.

Autoimmunity. A very important characteristic of the body's immune system is a capacity to distinguish between *self* and *nonself.* In terms of the immune system, the ability to make such distinctions proceeds at the biochemical level without conscious awareness of the individual. Less than a century ago, the ability of the body to distinguish self from nonself (foreign), at the biochemical level, was considered impossible. Over the years, however, much descriptive information accumulated which, without ample explanation, proved that the body does remember at the biochemical level. Seldom, for example, have medical records shown a second infection with mumps, measles, or smallpox, once an individual survived the first attack. The question, how does the body remember? remained a mystery for many decades until the concept of the immune system was first outlined in a very general way.

When the immune system functions property, which is the normal situation, antibodies to parts of the same body are not produced. But a condition known as *autoimmunity* can sometimes occur. In such circumstances, antibodies and sensitized (antigen-reactive) cells may be produced and directed against "self" antigens. To treat autoimmunity is one of the challenges of modern medicine. This autoimmunity mechanism may trigger some asthmatic paroxysms and is presently being considered as a suspect mechanism by cancer researchers—with considerable study being directed to tumor immunology. Some investigators hypothesize that the body system has the capacity to recognize neoplastic cells and to destroy them, but that in some individuals the process becomes inoperative, allowing the cells to multiply.

The ability of the body to tolerate self-antigens and thus to preclude autoimmunity is known as *immunologic tolerance*. As early as 1959, some scientists suggested that the antigen-specific lymphocytes which interact with self-antigens are eliminated during the prenatal state. In recent years, the concept of *suppressor mechanisms* has been well received. In this concept, the immune system, responding to an antigen, in addition to producing antibodies and sensitized cells to combat foreign substances, also initiates various suppressor mechanisms, causing some mediation of the process. Such substances are termed *mediators*. This may explain why, in some clinical situations, an early and strong immune response may be due to deficiencies in the suppressor mechanisms; or, in contrast, a deficient immune response may not be caused by the lack of particular lymphocytes, but rather by overreactive suppressor mechanisms.

A study of the body's immune system over the years has made possible the preparation of antitoxins and vaccines for preventing and treating many diseases. These studies have assisted in dealing with the problems of allergy and hypersensitivity and in finding partial or effective solutions to problems which arise from malperformance of the immune systems. Further progress during the 1980s is expected in the understanding and treatment of immune-system-related disorders, such as amyloidosis, macroglobulinemia, multiple myeloma, systemic lupus erythematosus, asthma, and various allergies, among others.

See also **Clone;** and **Immune System and Immunology.**

ANTICAKING AGENTS. Some products, particularly food products that contain one or more hygroscopic substances, require the addition of an *anticaking agent* to inhibit formation of aggregates and lumps and thus retain the free-flowing characteristic of the products. Calcium phosphate, for example, is commonly used in instant breakfast drinks and lemonade and other soft-drink mixes.

The general function of an anticaking agent can be described by using silica gel as an example. Generally anticaking agents are available as very small particles (ranging from 2 to 9 micrometers in diameter). A typical application for silica gel is admixture with orange-juice crystals to assure a free-flowing product, avoiding formation of crystal cakes and hard lumps. The very high adsorption properties of the anticaking substance removes moisture that can cause fusion. The billions of extremely fine, inert particles coat and separate each grain of powder (product) to keep it free-flowing. Many anticaking agents, including silica gel, also act as dispersants for powdered products. Many food products, when stirred into water, tend to form lumps which are difficult to disperse or dissolve. The agent not only improves flow properties, but also increases speed of dispersion by keeping the food particles separated and permitting water to wet them individually instead of forming lumps. As is true with so many food additive chemicals, anticaking agents serve multiple functions. In addition to acting as an anticaking and dispersing agent, silica gel also can be used as a moisture scavenger and carrier. Some additives, when they are capable of serving several functions, may be called *conditioning agents.*

Anticaking agents commonly used include: calcium carbonate, phosphate, silicate, and stearate; cellulose (microcrystalline); kaolin; magnesium carbonate, hydroxide, oxide, silicate, and stearate; myristates; palmitates; phosphates; silica (silicon dioxide); sodium ferrocyanide; sodium silicoaluminate; and starches.

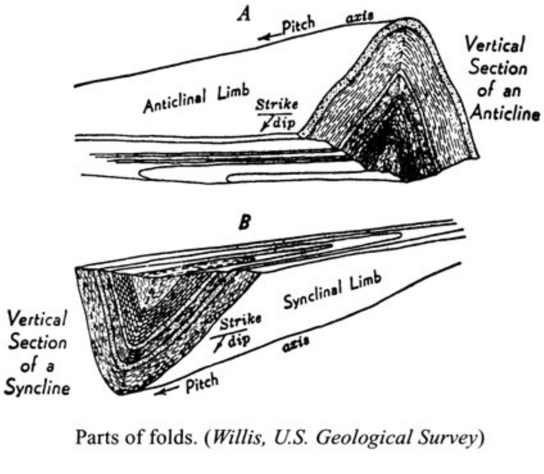
ANTICATHODE. In an x-ray tube, the target on which the electron beam is focused and from which the x-rays are radiated.

ANTICLINE. A folded structure involving bedded rocks in which the strata are arched upward so that the beds bend downward on either side. These downward-bending beds constitute the limbs of the fold.

The angle which the beds on the limbs of the fold make with the horizontal is spoken of as the dip. The term dip is also used to indicate the inclination of bedding in other structures.

Anticlinal arches may be broad and gentle or sharp with a steep dip, symmetrical or asymmetrical, or may be complicated by minor folds on the limbs. Anticlinal folds may be of sufficient magnitude to be measured in miles (kilometers), involving great thicknesses of sediments, or they may be so small as to be measured in inches (centimeters).

The direction of prolongation of the fold is termed the axis of the fold, and if not exactly horizontal the angle of inclination of the top bed of the anticline is called the pitch. Plunge is used as a synonym for pitch by some geologists.



Parts of folds. (*Willis, U.S. Geological Survey*)

ANTICLINORIUM. A composite anticlinal structure of folded beds is called an anticlinorium; a composite synclinal structure is called a synclinorium. The latter term, however, should be applied only to the compressed sedimentary filling of a geosyncline.

Section of an anticlinorium. (*Van Hise.*)

ANTICOAGULANTS. These are substances which prevent coagulation of the blood. For blood investigations made outside the body, sodium or potassium citrates, oxalates, and fluorides are sometimes used. For blood which is to be used for transfusions, sodium citrate is used.

Organic anticoagulants are used in vivo in the treatment of numerous conditions where blood coagulation can be dangerous, as in cerebral thrombosis and coronary heart disease, among others which will be described later. The main anticoagulants used are heparin and coumarin compounds, such as warfarin.

Heparin. A complex organic acid (mucopolysaccharide) present in mammalian tissues and a strong inhibitor of blood coagulation. Although the precise formula and structure of heparin are uncertain, it has been suggested that the formula for sodium heparinate, generally the form of the drug used in anticoagulant therapy, is $(C_{12}H_{16}NS_2Na_3)_{20}$ with a molecular weight of about 12,000. The commercial drug is derived from animal livers or lungs.

Heparin is considered a hazardous drug. Heparin may be the leading cause of drug-related deaths in hospitalized patients who are relatively well (Porter and Jick, 1978). It has been reported (Bell, et al, 1976) that some patients who receive continuously infused intravenous heparin develop *thrombocytopenia* (condition where the platelet count is less than 100,000 per cubic millimeter). Some authorities believe that the risk of thrombocytopenia associated with porcine heparin may be less than the risk associated with heparin of bovine origin.

Heparin, in addition to inhibiting reactions which lead to blood clotting, also inhibits the formation of fibrin clots, both in vitro and in vivo. Heparin acts at multiple sites in the normal coagulation system. Small amounts of heparin in combination with antithrombin III (heparin cofactor) can prevent the development of a hypercoagulable state by inactivating activated factor X, preventing the conversion of prothrombin to thrombin. Once a hypercoagulable state exists, larger doses of heparin, in combination with antithrombin II, can inhibit the coagulation process by inactivating thrombin and earlier clotting intermediates, thus preventing the conversion of fibrinogen to fibrin. Heparin also prevents the formation of a stable fibrin clot by inhibiting the activation of the fibrin stabilizing factor. The half-life of intravenously administered heparin is about 90 minutes.

Coumarin. Oral anticoagulants can be prepared from compounds with coumarin as a base. Coumarin has been known for well over a century and, in addition to its use pharmaceutically, it is also an excellent odor-enhancing agent. However, because of its toxicity, it is not permitted in food products in the United States (Food and Drug Administration). One commercial drug is 3-(alpha-acetonyl-4-nitrobenzyl)-4-hydroxycoumarin. This drug reduces the concentration of prothrombin in the blood and increases the prothrombin time by inhibiting the formation of prothrombin in the liver. The drug also interferes with the production of factors VII, IX, and X, so that their concentration in the blood is lowered during therapy. The inhibition of prothrombin involves interference with the action of vitamin K, and it has been postulated that the drug competes with vitamin K for an enzyme essential for prothrombin synthesis.

Another commercial drug is bis-hydroxycoumarin, $C_{19}H_{12}O_6$. The actions of this drug are similar to those just described.

Warfarin. This compound is also of the coumarin family. The formula is 3-(alpha-acetonylbenzyl)-4-hydroxycoumarin. In addition to use in anticoagulant therapy in medicine, the compound also has been used as a major ingredient in rodenticides, where the objective is to induce bleeding and, when used in heavy doses, is thus lethal. The compound can be prepared by the condensation of benzylideneacetone and 4-hydroxycoumarin.

The anticoagulant action of warfarin is through interference of the gamma-carboxylation of glutamic acid residues in the polypeptide chains of several of the vitamin K-dependent factors. The carboxylation reaction is required for the calcium-binding activity of the K-dependent factors. Because of the reserve of procoagulant proteins in the liver, usually several days are required to effect anticoagulation with warfarin.

Warfarin antagonists include vitamin K, barbiturates, gluthethimide, rifampin, and cholestyramine. Warfarin potentiators include phenylbutazone, oxyphenbutazone, anabolic steroids, clofibrate, aspirin, hepatotoxins, disulfiram, and metronidazole. In patients undergoing anticoagulation therapy with warfarin, it has been found that cimetidine (used in therapy of duodenal ulcer) may increase anticoagulant blood levels and consequently prolong the prothrombin time.

Anticoagulation Therapy

Prior to administration of anticoagulant drugs, patients must be carefully evaluated. Anticoagulant drugs are to be avoided if any of the following conditions prevail: a history of abnormal bleeding, recent corticosteroid therapy, recent intraocular or intracranial bleeding, recent pericarditis, and recent peptic ulcer or esophageal bleeding. A history of the individual's use of antiplatelet agents, such as aspirin, dipyridamole, phenylbutazone, and indomethacin, should be obtained and evaluated. Anticoagulant drugs should be administered with particular care during pregnancy. Heparin does not anticoagulate the fetus because it does not cross the placenta. Warfarin, on the other hand, anticoagulates both the mother and the fetus. Problems may arise from the administration of warfarin during pregnancy, particularly during the first trimester.

The bile sequestrant cholestyramine is frequently used in the treatment of familial hypercholesterolemia. This drug not only binds cholesterol, but also a number of other drugs, including anticoagulants.

Deep Vein Thrombosis and Pulmonary Embolism. Prompt administration of intravenous heparin is indicated in the treatment of this condition. Heparin is fast-acting, prevents further thrombus formation, and when used in therapeutic doses, also prevents the release of serotonin and thromboxane A_2 from platelets that adhere to thrombi that embolize to the lungs. The size of the dose required varies with a number of patient conditions. Several authorities are convinced that the continuous (pump-driven) infusion method is superior to intermittent injections.

Heparin is usually administered for a period ranging from 7 to 10 days. Frequently, during the last half of this period of heparin therapy, oral anticoagulation will be commenced with warfarin. The time during which oral anticoagulation administration should be continued may be three months or longer after clinical evidence that the venous thrombosis has subsided; and for one year after pulmonary embolism.

Cerebral Thrombosis. Among the specific modes of treatment that have been used in anticoagulation therapy. Many authorities suggest the use of anticoagulants for an evolving stroke in an effort to arrest the propagation of thrombus. In this procedure, a lumbar puncture is usually prepared first. If there is a presence of red blood cells in the spinal fluid, this infers a hemorrhagic infarction, in which case anticoagulants are withheld for a minimum of 48 hours. If the fluid is clear, heparin can be administered by continuous intravenous drip.

Cerebral Transient Ischemic Attack (TIA). Aspirin, as a platelet-inhibiting agent, has been found effective in the medical management of TIA. As studied by the Joint Committee for Stroke Facilities in the late 1970s, the results of anticoagulant therapy for TIA were reported as vague, but possibly this is due to poorly planned tests. However, anticoagulant therapy is considered a proper mode of treatment for persons who cannot tolerate aspirin, with warfarin the drug of choice. Oral anticoagulants are not given to persons with gastrointestinal ulcerations, severe hypertension, bleeding tendencies, or renal or hepatic failure.

Coronary Heart Disease. Long-term preventive anticoagulation with warfarin and similar drugs in patients with coronary artery disease has decreased in popularity during the last few years because evidence collected over a long period of time has not shown, in a convincing way, that the therapy is of value. As of the late 1980s, the therapy of choice includes the use of platelet-inhibiting drugs, notably acetylsalicylic acid (aspirin), sulfinpyrazone, and dipyridamole.

Prevention of Thromboembolism. Anticoagulation for prevention of thromboembolism is not used to the extent that it was once employed in the 1950s. Some professionals have reexamined the data of that period, however, and have concluded that anticoagulant therapy does have value, but their observations have not been widely accepted.

In present times, because of early mobilization and shorter stays in hospital, venous thrombosis in the legs and resulting pulmonary embolism has declined to a large degree. In persons with acute myocardial infarction, prophylactic low-dose heparin has reduced the incidence of venous thrombosis in the legs. It is considered as a reasonable alternative to warfarin in selected patients. Preventive anticoagulation may be indicated in some cases to prevent strokes due to left ventricular mitral thrombi embolizing in the brain.

Prosthetic Valve Endocarditis. Anticoagulants are sometimes used in the overall treatment of PVE even though there are risks of intracerebral hemorrhage or hemorrhagic infarction. Countering this risk, however, is the risk of major thromboembolic complications involving the central nervous system that may occur in the absence of continued anticoagulant therapy.

Anticoagulant therapy is also sometimes used in cases of congestive heart failure and in the treatment of polycythemia vera (elevation of the packed cell volume or the hemoglobin level) where not contraindicated.

Massive Venous Occlusion. This may be described as a surgical emergency. Immediately after diagnosis, intravenous heparinization is started and continued during the thrombectomy.

Mini-Dose Heparin. Small subcutaneous doses of heparin have been found to be effective in high-risk postsurgical patients and in patients with acute myocardial infarction. The preventive treatment is commenced a few hours before an operative procedure and continued postoperatively for 4 to 5 days. As the result of a study in 1975, low-dose heparin prophylaxis in high-risk patients who undergo abdomino-thoracic surgery has become a widely accepted practice. However, preventive anticoagulant therapy, to date, has been unsatisfactory and controversial in the instances of hip surgery or prostatectomy.

See also **Cerebrovascular Diseases; Heart and Circulatory System (Human); Ischemic Heart Disease.**

Additional Reading

Editor's Note: The following references have been selected to provide the interested reader with more detail on the pharmacologic complexities of anticoagulants. The article by Edwin W. Salzman, M.D. is a short, but excellent summary of the status of antithrombotic drugs as of early 1992. The article on heparin by Jack Hirsh, M.D. and the article on warfarin by the same author provide important fundamental backgkround information on the pharmacokinetics and pharmacodynamics of the most widely used anticoagulant drugs.

Brandjes, D. P. M.. et al.: "Acenocoumarol and Heparin Compared with Acenocoumarol Alone in the Initial Treatment of Proximal-Vein Thrombosis," *N. Eng. J. Med.*, 1485 (November 19, 1992).

Chesebro, J. H., Fuster, V, and J. L. Halperin: "Atrial Fibrillation — Risk Marker for Stroke," *N. Eng. J. Med.*, 1556 (November 29, 1990).

Cheesebro, J. H., and V. Foster: "Thrombosis in Unstable Angina," *N. Eng. J. Med.*, 192 (July 16, 1992).

Ezekowitz, M. D., et al.: "Warfarin in the Prevention of Stroke Associated with Nonrheumatic Atrial Fibrillation," *N. Eng. J. Med.*, 1406 (November 12, 1992).

Gold, H. K.: "Conjunctive Antithrombotic and Thrombolytic Therapy for Coronary-Artery Occlusion," *N. Eng. J. Med.*, 1483 (November 22, 1990).

Hirsh, J.: "Heparin," *N. Eng. J. Med.*, 1565 (May 30, 1991).

Hirsh, J.: "Drug Therapy: Oral Anticoagulant Drugs," *N. Eng. J. Med.*, 1865 (June 27, 1991).

Hull, R. D., et al.: "Heparin for 5 Days as Compared with 10 Days in the Initial Treatment of Proximal Venous Thrombosis," *N. Eng. J. Med.*, 1260 (May 3, 1990).

Hull, R. D., et al.: "Subcutaneous Low-Molecular-Weight Heparin Compared with Continuous Intravenous Heparin in the Treatment of Proximal-Vein Thrombosis," *N. Eng. J. Med.*, 975 (April 9, 1992).

Poller, L., and F. R. C. Path: "The Effect of Low-Dose Warfarin on on the Risk of Stroke in Patients with Nonrheumatic Atrial Fibrillation," *New Eng. J. Med.*, 129 (July 11, 1992).

Salzman, E. W.: "Low-Molecular-Weight Heparin and Other New Antithrombotic Drugs," *N. Eng. J. Med.*, 1017 (April 6, 1992).

Saour, J. N., et al.: "Trial of Different Intensities of Anticoagularion in Patients with Prosthetic Heart Valves," *N. Eng. J. Med.*, 428 (February 15, 1990).

Stroke Prevention in Atrial Fibrillation Study Group of Investigators: "Special Report," *N. Eng. J. Med.*, 863 (March 22, 1990).

Theroux, P.. et al.: "Reactivation of Unstable Angina after the Discontinuation of Heparin," *N. Eng. J. Med.*, 141 (July 16, 1992).

Thomas, D. P.: "Low-Molecular-Weight Heparin," *N. Eng. J. Med.*, 817 (September 1, 1992).

ANTICOINCIDENCE CIRCUIT. A circuit with two input terminals which delivers an output pulse if one input terminal receives a pulse, but delivers no output pulse if pulses are received by both input terminals simultaneously or within an assignable time interval.

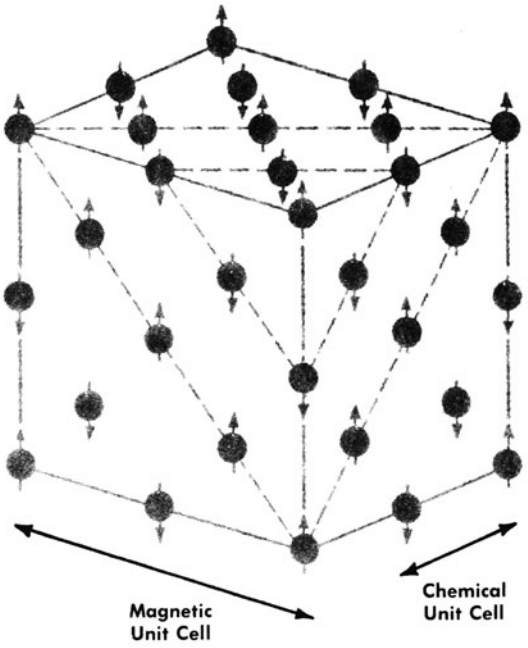
ANTICOINCIDENCE COUNTER. An arrangement of counters and associated circuits which will record a count if and only if an ionizing particle passes through certain of the counters but not through the others.

ANTICYCLONE. See **Atmosphere (Earth).**

ANTIDOTE. An agent which inhibits or counteracts the action of a poison. There is a wide variety of poisons, such as the *corrosives* (strong acids and alkalis) which cause local destruction of tissues; irritants which produce congestion of the organ with which they come in contact; the neurotoxins which affect the nerves or some of the basic processes within the cell; hemotoxins; hepatotoxins, and nephrotoxins. Consequently, the list of effective antidotes is long and reasonably complex, and usually much less lifesaving than making immediate efforts to have the poison victim vomit and thus expel as much of the poison as may be possible. Unless the appropriate antidote is selected, administration can be harmful rather than helpful. To illustrate this, if a sleep-producing drug, such as opium or morphine, has been taken in overdosage, it is best to keep the patient awake by giving strong coffee. In contrast, in the instance of strychnine poisoning, no stimulants should be given and the patient should be kept as quiet as possible. In every type of poisoning, immediate medical aid is essential. Most local health departments have lists of antidotes for common poisons.

ANTIFERROELECTRIC. Certain crystals, such as tungstic oxide WO_3, ammonium dihydrogen phosphate $(NH_4)H_2PO_4$, sodium niobate $NaNbO_3$, and disilver trihydrogen paraperiodate $Ag_2H_3IO_6$ have been shown to exhibit spontaneous microscopic polarization similar to that in ferroelectrics, except that different types of ions are polarized in different directions, so that the megascopic spontaneous polarization is small, or even vanishing. These materials are described as being antiferroelectric. Strictly the term is applied only to substances in which the net spontaneous polarization is zero; materials in which the polarizations of the individual ions cancel only in part are often referred to as *quasi-ferroelectrics* or as *ferrielectrics*. The relation of antiferroelectrics to ferroelectrics is similar to that of antiferromagnetics to ferromagnetics. See **Antiferromagnetism; Ferromagnetism.**

ANTIFERROMAGNETISM. The observed susceptibility curves of certain substances suggest that the system has gone into a state analogous to the ferromagnetic state, but with neighboring spins antiparallel, instead of parallel. See accompanying figure. That is, such substances exhibit a paramagnetism (low positive susceptibility) that varies with temperature in a manner similar to ferromagnetism, exhibiting a Curie point. Their resulting superlattices have been observed by neutron diffraction. The interaction giving preference to the antiparallel arrangement is believed to be an exchange force, similar to that invoked in the Heisenberg theory of ferromagnetism, but opposite in sign. Evidence from face-centered crystals has suggested the importance of super-exchange between next-nearest neighbors, through the anions. The alignment of the ions can be removed by heating the crystal, and the tem-



Antiferromagnetic state spin structure of manganese oxide as determined by Shull, Strauser, and Woolan.

perature at which the ordered spin arrangement breaks down is called the *Néel temperature*.

ANTIGEN. A substance, usually a protein, a polysaccharide, or a lipoid, which when introduced into the body stimulates the production of antibodies. Bacteria, their toxins, red blood corpuscles, tissue extracts, pollens, dust, and many other substances may act as antigens. See **Antibody.**

ANTIHISTAMINE. A synthetic substance essentially structurally analogous to histamine, the presence of which in minute amounts prevents or counteracts the action of excess histamine formed in body tissues. See also **Histamine.** Antihistamines are usually complex amines of various types. They find a number of medical uses.

In immediate hypersensitivity situations (reaction between antigen and antibody as encountered in hay fever, hives (urticaria), allergic (extrinsic) asthma, bites, drug injections, among others), antihistamines can be part of the effective therapy. Although widely used, antihistamines and steroids are not always the drugs of choice. In atopic dermatitis (chronic skin disorder), antihistamines may assist in breaking the itch-scratch cycle, particularly in persons whose sleep may be interrupted by pruritus. Antihistamine compounds for urticaria are also effective for atopic dermatitis therapy. Frequently, shifting from one antihistamine to another is effective and helps to reduce side effects of the drugs. Antihistamines are sometimes effective in the treatment of autoerythroyce purpura, a rare disease. Antihistamines are also used in connection with mild penicillin reactions, and in cases of penicillin desensitization procedures. Certain antihistamines find application to control mild parkinsonism.

Some antihistamines are particularly effective in alleviating the onset of motion sickness. Some antihistamines have been found helpful in relieving persistent, unproductive coughs that frequently accompany bronchitis or coughs associated with allergy. They are used in connection with perennial and seasonal allergic rhinitis.

Most antihistamines have anticholinergic (drying) and sedative side effects, sometimes producing marked drowsiness and reduction of mental alertness and thus should not be used by persons who operate machinery, drive vehicles, or otherwise must react quickly. Because of their similar structure, antihistamines appear to compete with histamine for cell receptor sites. Although conventional antihistaminic drugs, such as mepyramine, block the allergic and smooth muscle effects caused by histamine, the structure of these drugs is not sufficiently similar to histamine to inhibit histamine-stimulated gastric acid secretion. However, during the last few years, so-called histamine-blocking drugs have been developed which appear to be effective. It has been found that such compounds must contain the imidazole ring of histamine, with their potency enhanced by extension of the side chain. Among these new drugs are metiamide and cimetidine.

Some drugs in the antihistamine series play markedly different roles. Hydroxyzine hydrochloride and hydroxyzine pamoate have been used in the total management of anxiety, tension, and psychomotor agitation in conditions of emotional stress, usually requiring a combined approach of psychotherapy and chemotherapy. Hydroxyzine has been found to be particularly useful for making the disturbed patient more amenable to psychotherapy in long-term treatment of the psychoneurotic and the psychotic. The drug is not used as the only treatment of psychosis or of clearly demonstrated cases of depression. Hydroxyzine has also been found useful in alleviating the manifestations of anxiety and tension in acute emotional problems and in such situations as preparation for dental procedures. Hydroxyzine therapy has been used in treatment of chronic alcoholism where anxiety withdrawal symptoms or delirium tremens may be present. Hydroxyzine may potentiate narcotics and barbiturates.

Most conventional antihistamines are available for both oral and intravenous or intramuscular administration. In serious cases of urticaria (hives), for example, the injection rather than oral route is most effective. The major excretion route for most antihistamines is hepatic (liver), occurring within 4 to 15 hours.

In addition to the side effects previously mentioned, some antihistamines may cause neutropenia (neutrophil count in the blood is less than 1800 per cubic millimeter). Some antihistamines also may cause a modification of normal platelets in the blood.

Some of the more commonly used antihistamine compounds are listed below:

- Ethanolamines:
  - Diphenhydramine hydrochloride (Benadryl®)
  - Dimenhydrinate (Dramamine®)
- Ethylenediamines:
  - Tripelennamine hydrochloride (Pyribenzamine®)
- Alkylamines:
  - Chlorpheniramine maleate (Chlor-Trimeton®)
- Piperazines:
  - Cyclizine hydrochloride (Marezine®)
- Phenothiazines:
  - Promethazine hydrochloride (Phenergan®)
- Others:
  - Cyproheptadine hydrochloride (Periactin®)
  - Hydroxyzine hydrochloride (Atarax®)
  - Hydroxyzine pamoate (Vistaril®)

ANTIHUNTING CIRCUIT. A stabilizing or equalizing circuit used in a closed-loop feedback system to modify the response of the system in order that self-oscillations may be prevented.

An *antihunting transformer* is sometimes used in dc feedback systems as a stabilizing network. The primary of this transformer is in series with the load connected to the system. The secondary of the transformer has a voltage which is proportional to the derivative of the primary current, and is thus an appropriate signal to be fed back into some other part of the loop to prevent self-oscillations.

ANTILLES CURRENT. An ocean current, the northern branch of the north equatorial current flowing along the northern side of the Great Antilles carrying water that is identical with that of the Sargasso Sea. The Antilles current eventually joins the Florida current (after the latter emerges from the Straits of Florida) to form the Gulf Stream.

ANTIMATTER. Matter that consists of antiparticles. One of the great discoveries of modern physics was that for every type of elementary entity of matter and radiation (particle), there exists a corresponding conjugate type of entity (an antiparticle). In the antiparticle, certain of the particle-defining properties are identical (*conjugation invariant*), and others are reversed in sign (*conjugation reversing*). The reversed sign in a conjugation-reversing property allows one to maintain a conservation law for that property in the dramatic processes of *pair creation* and *pair annihilation* in which an antiparticle is observed to appear and disappear together with the particle to which it is conjugate. In those cases where all the conjugation-reversing properties occur with zero values, the antiparticle is identical with the particle. The progressive recognition of the existence of antiparticles was initiated by Dirac's relativistic antielectron theory in 1931, and by Anderson's independent experimental discovery of the antielectron (positron) in 1932.

Particles and (antiparticles) include: Electron (positron); Proton (anti-proton); Electron neutrino (antielectron neutrino); Muon neutrino (anti-muon neutrino); Neutron (antineutron); Positive pion (negative pion), etc.

The principle of charge conjugation symmetry states that if each particle in a given system is replaced by its corresponding antiparticle, then an observer will be unable to tell the difference. For example, if in a hydrogen atom, the proton is replaced by an antiproton and the electron is replaced by a positron, then this antimatter atom, if observed by persons also made of antimatter, will behave exactly like an ordinary atom. In an antimatter universe, the laws of nature could not be distinguished from the laws of an ordinary matter universe. However, it turns out that there are certain types of reactions where this rule does not hold, and these are just the types of reactions where conservation of parity breaks down. For an explanation of this, see long footnote in early portion of entry on **Particles (Subatomic).** Also see list of related entries at end of aforementioned entry.

ANTIMER. See **Amino Acids.**

ANTIMETABOLITES. These substances fall into the general class of cytotoxic chemicals, i.e., agents that damage cells to which they are applied. Antimetabolites are so similar to normal enzymatic substrate molecules or metabolites as to gain entry into the cellular machinery of intermediary metabolism, but once there they differ enough to cause enzymatic inhibition. If incoporated into protein, nucleic acids, or coenzymes, for example, they will diminish the biological worth of those substances. Spectacular agents of this sort include the antifolic acids, such as aminopterin and amethopterin, various other vitamin analogs, and analogs of the naturally occurring purines, pyrimidines, nucleosides, and amino acids. Effective action against the integrity of the cell appears to be exerted at a number of points of intermediary metabolism in these multifarious antimetabolites. Of particular interest with many of them is an interference in normal nucleic acid metabolism. 5-Fluoro-2'-deoxyuridine, for example, acts to inhibit the synthesis of thymidylate, a necessary precursor of DNA, and the related 5-bromo-2'-deoxyuridine is actually incorporated into new DNA in the place of thymidine. Both of these agents increase the frequency of chromosomal disturbances. Various other base analogs, if incorporated into DNA, can lead to gene mutation by alteration of the normal sequence of nucleotides during replication through incorrect base pairing. 2-Aminopurine is an example of such a mutagen. 8-Azaguanine can be incorporated into ribonucleic acids, which are thus rendered defective. Among the actions of 6-mercaptopurine is an interference in the biochemical activity of coenzyme A, with resultant mitochondrial damage. Such amino acid analogs as β-fluorophenylalanine can effectively halt cellular activities by being incorporated into new proteins, which thereupon fail to attain their proper enzymatic or other functions.

Advantage is taken of the properties of antimetabolites in chemotherapy. In cancer chemotherapy, several antimetabolites are used. These include methotrexate, 6-mercaptopurine, 6-thioguanine, 5-fluorouracil, and cystine arabinoside. In the chemotherapy of metastatic breast cancer, 5-fluorouracil and methotrexate, in combination with cyclophosphamide, have been used. Antimetabolites, sometimes along with corticosteroids, are used in the therapy of various autoimmune diseases, such as thrombocytenic purpura, thyroiditis, Goodpasture's syndrome, among others.

Metabolites are implicated as agents that produce marrow aplasia as found in leukemia.

Additional Reading

Note: Check references listed in articles on **Cancer; Cancer Research; Gene Science;** and **Industrial Biotechnology**.

ANTIMICROBIAL AGENTS (Foods). Frequently substances are added to food products or applied to the surface of some foods while in transit or storage for the purpose of inhibiting the growth of certain destructive microorganisms. Such organisms affect a large percentage of fresh fruits and vegetables, as well as meat that has been cut and packaged. In addition to spoilage of the foods themselves, inadvertent consumption of certain molds can cause human disease. See also **Foodborne Diseases.**

Antimicrobials for foods act against microorganisms by (1) adversely affecting the cellular membranes of the destructive molds, yeasts, bacteria, etc.; (2) by interfering with the genetic mechanisms of the offending microorganisms; and (3) by interfering with cellular membranes of the microorganisms. The chemicals used for such agents function against the microorganisms not unlike the actions of antibiotics and other antimicrobial drugs in preventing or arresting microbially caused infections in humans and other animals.

For several years, the use of antimicrobial agents in connection with food products has been strictly regulated in the United States and most of the other developed countries. Consequently, this is a field that is subject to constant change. This is evidenced by continuing controversy concerning, for example, the use of nitrates and nitrites and, more recently, the use of sulfur dioxide and sulfites. In the United States, the best source for current regulations is the Food and Drug Administration, Washington, D.C.

The principal antimicrobial agents still permitted, in most countries, for use in foods are benzoic acid and sodium benzoate; the parabens; sorbic acid and sorbates; propionic acid and propionates; acetic acid and acetates; nitrates and nitrites; sulfur dioxide and sulfites; diethyl pyrocarbonate; epoxides; hydrogen peroxide; and phosphates. These agents are not used uniformly—some are applied to the food products near point of sale or consumption; others are introduced directly into the products during processing; still others are used only in connection with controlling microorganisms that may collect in the processing equipment. For the latter purpose, inasmuch as the equipment is thoroughly washed after control chemicals are used, many stronger and toxic (to humans in high concentrations) can be used, including hypochlorites and strong detergents. Because the need to maintain food processing equipment in ultraclean condition at all times, microbial decontamination is practiced on a shift or daily basis. To alleviate many of the problems and loss of time required by such cleaning operations, much food processing equipment has been designed for cleaning in place (CIP), where the cleaning is essentially automatic and does not require disassembly and reassembly of critical parts. Antimicrobial agents are also used in the packaging and wrapping of a number of food products, notably in connection with certain fruits and vegetables. Numerous studies have been made and, consequently, regulations established concerning the possible migration of such control chemicals from the packaging materials to the consumable product. Also, very powerful antimicrobial agents are permitted and used for special purposes, such as fumigating grain elevators and bulk food storage warehouses, where the objective is not simply that of killing microorganisms, but also to kill insects, rodents, and the like. See article on **Grain Storage Insects.**

Although food irradiation has been proved effective and could ultimately significantly alter the need for antimicrobial chemical agents, there remain a number of regulatory as well as consumer acception questions that require further resolution. This topic is discussed in detail in the "Foods and Food Production Encyclopedia," (D. M. and G. D. Considine, Eds.), Van Nostrand Reinhold, New York, 1982.

Benzoic Acid and Sodium Benzoate. These compounds are most active against yeasts and are less effective against molds. They are best suited for foods with a natural or adjusted pH below 4.5. The average dosage in foods ranges between 0.05–0.1% (weight), depending upon product. Benzoic acid occurs naturally in cinnamon, ripe cloves, cranberries, greengage plums, and prunes. See also **Benzoic Acid.** For a number of years the sodium salt has been preferred over the acid by food processors.

Common applications for these compounds include carbonated and noncarbonated beverages, but excluding beers and wines because of their action against yeasts. They are also used in salted margarines, jams, jellies, and preserves, pie fillings, salads and salad dressings, pickles, relishes, and other condiments, as well as olives and sauerkraut. In terms of human metabolism of these substances, some authorities have suggested that benzoate is conjugated with glycine to produce hippuric acid, which is excreted, possibly accounting for 65–95% of benzoate ingested. It has been postulated that the remainder is detoxified by conjugation with glycuronic acid.

Parabens. These compounds include the methyl, ethyl, propyl, and butyl esters of para-hydroxybenzoic acid. In the United States and a number of other countries, the methyl and propyl esters are preferred, while European food processors favor the ethyl and butyl esters. The parabens were first described in 1924 as having antimicrobial activity and initially were used in cosmetic and pharmaceutical products.

The parabens are most effective against molds and yeasts, but less active against bacteria, particularly gram-negative bacteria. The antimicrobial activity of the parabens is directly related to the molecular chain length (methyl is weakest; butyl is strongest). However, the solubility of these compounds is in inverse relationship with chain length. These characteristics give rise to the use of two or more esters in combination and sometimes in combination with entirely different antimicrobial agents, such as sodium benzoate. Below a pH of 7, the parabens are only weakly effective.

The parabens are used in carbonated beverages and other soft drinks, including cider. In lieu of pasteurizing or using Millipore filtration, some brewers use the parabens for controlling secondary yeast formation. Because of their activity against yeasts, they are not used in bread and rolls, but they find use in other bakery products, such as pie crusts, certain pastries, icings, toppings, fillings, and cakes. The parabens are particularly effective in preserving fruit cakes. Usage also includes creams and pastes, fruit products, flavor extracts, pickles and olives, and artificially sweetened jams, jellies, and preserves. Average dosage ranges from 0.03–0.6% (weight).

Sorbic Acid and Sorbates. Sorbic acid and its potassium and sodium salts are effective against molds, but less effective against bacteria. These compounds may be incorporated directly into the food product, but they are frequently applied by spraying, dipping, or coating. The compounds are effective up to a pH of about 6.5. This is higher than propionates and sodium benzoate, but not so high as the parabens. Metabolism in humans parallels that of other fatty acids.

Because sorbates affect yeasts, the compounds are not directly useful in yeast-raised goods. Sorbates are particularly favored for use in chocolate syrups. They can be used in wine production in conjunction with sulfur dioxide against bacteria and are effective in inhibiting development of unwanted yeasts. They are also used in artifically sweetened jellies, jams, and preserves; in pickles and related products; in nonsalted margarines; in dried and smoked fish products; in semimoist pet foods; in dry sausage castings; in fruit-filled toaster pastries; and in cheese and cheese products. In the latter products, the agents usually are applied by dipping or spraying. Wrappers also may be impregnated with sorbates.

Propionic Acid and Propionates. The antimicrobial properties of propionic acid and its calcium and sodium salts were first noted in 1913. Today, the calcium and sodium salts are most commonly used. These compounds are more active against molds than sodium benzoate, but have little if any activity against yeasts. The propionates are well known for their effectiveness against *Bacillus messentericus*, a "rope"-forming microorganism. These compounds are effective up to a pH of 5 or slightly higher. Metabolism in the human body parallels that of other fatty acids.

An early application for the propionates was that of dipping cheddar cheese in an 8% propionic acid solution. This increased mold-free life by 4 to 5 times more than when no preservative was added. For pasteurized process cheese and cheese products, propionates can be added before or with emulsifying salts. Research has indicated that propionate-treated parchment wrappers provide protection for butter.

Use of propionates in breads can extend mold-free life by 8 days or more. Propionates are favored by bakers because of their effectiveness against ropy mold in breads up to pH levels of 6. For cakes and unleavened bakery goods, the sodium salt is usually preferred; for bread, the calcium salt is favored. This additive also contributes to the mineral enrichment of the product.

Acetic Acid and Acetates. Acetic acid (pure and as vinegar) and calcium, potassium, and sodium acetates, as well as sodium diacetate, serve as antimicrobial agents. In the United States, vinegar can contain no less than 4 grams of acetic acid per 100 milliliters of product. Acetic acid and calcium acetate are most effective against yeasts and bacteria, and to a lesser extent, molds. The diacetate is effective against both rope and mold in bread. It is interesting to note that the antimicrobial effectiveness of acetic acid and its salts is increased as the pH is lowered.

Optimal pH range varies with products and target microorganisms, but generally falls between 3.5 and 5.5. These agents are particularly effective against *Salmonella aertrycke. Staphylococcus aureus, Phytomonas phaseoli, Bacillus cereus, B. mesetericut. Saccharomyces cerevisiae,* and *Aspergillus niger*.

Unfortunately, to be effective against microorganisms in bakery products, acetic acid concentrations must be so high that an overly sour taste is imparted to the products. Sodium diacetate, however, can be used in small concentrations in bread and rolls to control rope and molds. Traditional concentrations of the acetate are 0.4 part to 100 parts of flour. During recent years, the propionates have largely displaced sodium diacetate for this use.

Vinegar or acetic acid is used in a number of products as much for its sour taste as for its antimicrobial properties. Such products include catsup, mayonnaise, pickles, salad dressing, and various condiment sauces. These agents also have been used to a lesser extent in malt syrups and concentrates, cheeses, and in the treatment of parchment wrappers for products, such as butter, to inhibit mold.

Nitrates and Nitrites. For many decades, sodium nitrate and nitrite, and potassium nitrate and nitrite have been used to cure, preserve, and provide a characteristic flavor to such meats as bacon, corned beef, frankfurters, ham, and various sausages. This tradition continues into the late 1980s, but was seriously threatened in the mid-and late 1970s. Some researchers reported that N-nitrosopyrolidine (NPyr) formed in bacon upon application of heat during preparation for consumption. It was observed that there was a greater concentration of the NPyr in adipose tissue than in the lean portion. A connection was proposed that involved serious implications of the ultimate carcinogenic risk involved in meat treated with the nitrates and nitrites. Numerous tests proceeded. One of the main factors learned during this period was how little knowledge food scientists had concerning the fate of the nitrates and nitrites. Although the precursors of the nitrosamines formed under certain conditions of cooking were known, the mechanism of formation was unknown.

Sulfur Dioxide and Sulfites. The use of sulfur dioxide gas and with it the production of sulfites differ somewhat from the other antimicrobial agents thus far described. Historical records show that burning sulfur to produce sulfur dioxide (SO_2 gas) dates back to the ancient Egyptians and Romans who used it in connection with wine making. See **Sulfur.** Action by sulfur dioxide is accomplished in the gaseous phase. The effect of SO_2 is markedly determined by concentration and pH conditions of the target product. Research has demonstrated that most bacteria are inhibited by HSO_3^- at concentrations of 200 parts per million (ppm) or less. With few exceptions, yeasts are also similarly inhibited. There are, however, some strains of molds that are considerably more resistant. The sulfite salts tend to be unstable and oxidize during long periods of storage, thus decreasing the availability of SO_2. This process is aggravated by the presence of moisture.

The most effective range for optimal microbial inhibition with sulfites is a pH of 2.5 to 3. It has been found that from 2 to 4 times greater concentrations of SO_2 are needed to inhibit the growth of microorganisms at a pH of 3.5 than at a pH of 2.5. It also has been demonstrated that, at a pH of 7, SO_2 has little if any effect on yeasts and molds, even at concentrations up to 1000 ppm. The inhibitory effects against bacteria are also considerably less when pH rises above 3.5. Some researchers believe that at higher pH levels, penetration of cell walls is much more difficult.

Because residual levels of sulfites in excess of 500 ppm impart a noticeable taste to food substances, this fact, regardless of any regulations toward limiting concentration, requires SO_2 levels to be controlled.

The use of SO_2 for preserving fruit juices, syrups, concentrates, and purees is particularly attractive in regions with warm climates and where products must be stored in bulk prior to processing. In these situations, the SO_2 concentration will range between 350 and 600 ppm. High sugar concentrations require higher levels of SO_2. For optimal effectiveness, the pH of some products has to be reduced.

It is a common practice to expose many fruits to SO_2 prior to dehydration. The SO_2 also extends storage life of raw fruit prior to dehydration. The optimal temperature for exposure to the gas is from 43 to 49°C. Unlike fruit, vegetables are usually dipped in solutions of neutral sulfites and bisulfites. Suggested levels of SO_2 in some dried and dehydrated fruits and vegetables, in parts per million, are:

Apricots, peaches, and nectarines	2000
Raisins	800 to 1500
Nectarines	2000
Pears	1000
Apples	800
Cabbage	750 to 1000
Carrots and potatoes	200 to 250

It is important to note, however, that bulk-treated fruits intended for canning should not have a residual level in excess of 20 ppm SO_2 because of possible sulfide (black precipitate) forming in the can as the result of hydrogen sulfide generation.

Many countries do not allow use of SO_2 or sulfite salts on meats, fish, processed meat and fish products, or fresh fruits and vegetables. Where permitted, sulfite is helpful in eliminating "black spot" formation in shrimp. End-use application of sulfite related compounds as, for example, to prevent discoloration of leafy vegetables on salad bars, is now subject to regulation in the United States.

A major use of sulfites is in wine making. It is used for sanitizing equipment and, prior to fermenting, the grape *musts* have to be treated with sulfites to inhibit the growth of any natural microbial flora present. This is done prior to the addition of pure cultures of the appropriate wine-making yeasts. While fermenting SO_2 also can function as an anti-oxidant, clarifier, and dissolving agent. Sulfur dioxide is often used after fermentation to prevent undesirable postfermentation alterations by various microorganisms. Levels of SO_2 during fermentation range from 50 to 100 ppm, depending upon condition of the grapes, temperature, pH, and sugar concentration. The wine industry uses sulfur dioxide dissolved in water, vaporized SO_2, and sulfite salts. An SO_2 level of 50–75 ppm assists the prevention of bacterial spoilage during the bulk storage of wine after fermentation.

Antibiotics. Much attention has been given to the use of antibiotics in food-associated applications since the introduction of penicillin in the 1940s. Their use has been limited. The use of antibiotics in animal feedstuffs continues to remain a topic of controversy in many countries; in other countries they have been banned. Antibiotics carry into the meat produced and further into human diets, thus possibly reducing their effectiveness in the treatment of human diseases.

Diethyl Pyrocarbonate. The preservative qualities of this compound were not recognized until the late 1930s and research on the compound continues to date. Also called pyrocarbonic acid diethyl ester, the compound is extremely effective against yeasts. It is also active against bacteria, such as *Lactobacillus pastorianus*, and various molds. The substance is generally used in still wines, fermented malt beverages, and noncarbonated soft drinks, as well as fruit-based beverages. Regulations on its use vary from one country to the next. Effective inhibition by diethyl pyrocarbonate is largely confined to acid products of low microorganism count. Some researchers point out that the pH should be less than 4 and that the microorganism count should not exceed 500 per milliliter. Some authorities observe that because of its rapid hydrolysis, no toxicity or residue problems should occur in products where the compound is permitted.

Epoxides. Two compounds are included in this category of antimicrobials. One is the gas, *ethylene oxide*; the other, *propylene oxide*, a colorless liquid with a boiling point of 35°C. Ethylene is highly reactive and must be used carefully and only with proper equipment. Somewhat less hazardous from an explosion standpoint, propylene oxide also has an explosive range of 2–22%. Consequently, these materials are usually mixed with inert substances, such as carbon dioxide or organic diluents.

Ethylene oxide is a universal antimicrobial in that it is lethal to all microorganisms. However, it is not universal from the standpoint of application. Propylene oxide is considered a broad-range microbiocide. In practically all aspects, propylene oxide is a considerably less effective agent, requiring longer exposures and greater concentrations because of its low penetrating power. However, propylene oxide is less toxic to humans.

The use of these gases (propylene oxide is volatilized) has been called "cold sterilization" and is frequently useful in sterilizing a number of low-moisture ingredients which end up on high-moisture foods. This prior sterilization lessens the total load on later thermal processing. The ability of these gases to kill microorganisms in low-moisture foods is an outstanding advantage. At the same time, macroorganisms also are killed. The gases find application in connection with spices, starches, nut meats, dried prunes, and glacé fruit. They are not used on peanuts (groundnuts).

Hydrogen Peroxide. Although usually regarded as a bleaching and oxidizing agent, this compound, H_2O_2, can be an effective antimicrobial and can be particularly useful in sterilizing processing equipment and packaging materials, notably prior to the aseptic packaging process. Regulations regarding the use of hydrogen peroxide vary from one country to the next.

Phosphates. The various phosphates are an effective multipurpose food additive chemical and functions other than their antimicrobial properties are usually given the greatest stress. The antimicrobial properties of the phosphates have been investigated over the years and are reasonably well documented.

Additional Reading

Beuchat, L. R.. and D. A. Golden: "Antimicrobials Occurring Naturally in Foods," *Food Technology*, 134 (January 1989).

Daeschel, M. A.: "Antimicrobial Substances from Lactic Acid Bacteria for Use as Food Preservatives," *Food Technology*, 164 (January 1989)

Davidson, P. M.. and M. E. Parish: "Methods for Testing the Efficiency of Food Antimicrobials," *Food Technology*, 148 (January 1989).

Jacoby, G. A., and G. L. Archer: "New Mechanisms of Bacterial Resistance to Antimicrobial Agents," *N. Eng. J. Med.*, 601 (February 28, 1991).

Roberts, T. A.: "Combination of Antimicrobials and Processing Methods," *Food Technology*, 156 (January 1989).

Wagner, M. K.. and L. J. Moberg: "Present and Future Use of Traditional Antimicrobials," *Food Technology*, 143 (January 1989).

ANTIMONY. Chemical element, symbol Sb, at. no. 51, at. wt. 121.75, periodic table group 15, mp 630.5°C, bp 1950°C, sp gr 6.62 (vacuum-distilled solid at 20°C) and 6.73 (single crystal). Naturally occurring isotopes are ^{121}Sb and ^{123}Sb. Antimony metal is a lustrous, silvery, blue-white solid, extremely brittle, and exhibiting a scalelike or flaky crystalline texture. The metal is easy to pulverize. The pure metal has a hardness of 3.0–3.3 on the Mohs scale and 55 on the Brinell scale. Of the more common metals, antimony is the poorest conductor (4.5 on a scale of 100 for copper). From careful studies, it has been observed that Sb contracts upon solidification rather than expanding. The element was first described by Thölden (Valentine) in 1450.

There are two natural isotopes, ^{121}Sb and ^{123}Sb; and ten radioactive isotopes, ^{116}Sb through ^{120}Sb, ^{122}Sb, and ^{124}Sb through ^{127}Sb. ^{124}Sb is used as a radiation source in industrial instruments for the measurement of flow of slurries and interface measurements in pipelines. See also **Radioactivity.**

First ionization potential 8.64 eV; second 16.5 eV; third 25.3 eV; fourth 44.1 eV; fifth 56 eV. Oxidation potentials $Sb + H_2O \rightarrow SbO^+ + 2H^+ + 3e^-$, -0.212 V,

$$2SbO^+ + 3H_2O \rightarrow Sb_2O_5 + 6H^+ + 4e^-,$$

-0.581 V, $Sb + 4OH^- \rightarrow SbO_2^- + 2H_2O + 3e^-$, 0.66 V. Other important physical properties of antimony are given under **Chemical Elements.**

Antimony exists in a number of allotropic forms. Gray or metallic antimony, density 6.79 g/per cm^3, is the stable form, forming rhombohedral crystals. Its vapor is that of Sb_4 up to 800°C, where dissociation to Sb_2 commences. Yellow antimony, Sb_4, density 5.3 g/cm^3 is less stable than yellow arsenic. It is produced by oxidation of stibine (see below) at very low temperatures, above which it is unstable. It changes even in the dark to black antimony at −90°C (in the light at −180°C). Black antimony, produced most readily by cooling antimony vapor or oxidizing stibine at 40°C, density 5.3, is metastable with respect to the gray form. It is also more reactive, igniting in air at room temperatures or above. Explosive antimony is produced by rapid electrodeposition of antimony from its halides. When heated or scratched, it undergoes an exothermic transformation to gray antimony. Its structure is amorphous, and differs somewhat from that of gray antimony.

Antimony is used in alloys, with lead for storage battery plates, with lead and tin in type metals and body solders, with tin and copper in bearing or antifriction metals. Antimony occurs chiefly as the sulfide (stibnite, Sb_2S_3) which is produced mainly in China, only small amounts in Mexico and Bolivia. Stibnite is (1) melted and reduced to antimony by iron metal and separated from fused ferrous sulfide (See also **Stibnite**); (2) roasted in air, and sublimed antimonous oxide collected and reduced by heating to fusion with carbon and sodium carbonate.

Antimony is also leached from tetrahedryte ore and recovered by electrowinning.

Antimony is scarcely tarnished in dry air but oxidized slowly in moist air; burns at a red heat in air or oxygen with incandescence forming antimonous oxide; insoluble in HCl; converted by HNO_3 into antimonous oxide or antimonic oxide, depending upon the concentration of acid; by chlorine into trichloride or pentachloride, by NaOH solution into antimonite.

Stibine: SbH_3, is formed by hydrolysis of some metal antimonides or reduction (with hydrogen produced by addition of zinc and HCl) of antimony compounds, as in the Gutzeit test. It is decomposed by aqueous bases, in contrast with arsine. It reacts with metals at higher temperatures to give the antimonides. The antimonides of elements of group 1a, 2a, and 3a usually are stoichiometric, with antimony trivalent. With other metals, the binary compounds are essentially intermetallic, with such exceptions as the nickel series, Ni_2Sb_3, NiSb, Ni_5Sb_2 and Ni_4Sb.

Trihalides: SbF_3, $SbCl_3$, $SbBr_3$, and SbI_3, are solids, and have pyramidal structures. Except for the fluoride, which is not hydrolyzed, they undergo partial hydrolysis only (in contrast with the phosphorus trihalides) on contact with water to yield insoluble oxyhalides, either of composition SbOX or varying somewhat from this composition to give such compounds as $Sb_4O_5Cl_2$. The antimony pentahalides, SbF_5 and $SbCl_5$ can be prepared, but the pentabromide exists only in double compounds, known as bromoantimonates, those for monovalent metals being of the type $MSbBr_6$, plus water of hydration, and yielding $SbBr_6^-$ ions. $SbCl_6^-$ and SbF_6^- ions are also known. Mixture of antimony(III) chloride, $SbCl_3$, in HCl solution with antimony(V) chloride, $SbCl_5$, in equimolar proportions yields a dark-colored solution. While antimony(IV) chloride cannot be isolated from it, compounds such as cesium antimony(IV) chloride, Cs_2SbCl_6 are formed by addition of cesium chloride, CsCl, and they are isomorphous with similar compounds of lead, tin, and other metals. However, tetravalent antimony should be paramagnetic because of the unpaired electron, whereas compounds of $SbCl_6^{2-}$are diamagnetic. Therefore it may be that these compounds contain equimolar mixtures of $SbCl_6^-$ and $SbCl_6^{-3}$. The existence of these higher halide complexes with tin (and bismuth), but not with phosphorus or arsenic, may be due to steric considerations.

Antimony(III) oxide: Sb_2O_3 or Sb_4O_6, is formed by melting antimony in air, or from the hydroxide $Sb(OH)_3$. The Sb_2O_3 of commerce is produced from the oxidation of stibnite ore. Antimony is below arsenic in the periodic table, and $Sb(OH)_3$ is more definitely amphiprotic than $As(OH)_3$, forming not only antimony(III) salts and antimonites (containing the ion SbO_2^- or $Sb(OH)_4^-$), but also basic salts, especially the antimonyl salts, containing the ion SbO^+. Antimony(V) oxide, formed by oxidation of the metal with HNO_3, is less soluble in H_2O than As_2O_5. Antimonic acid cannot be obtained by hydration, and the product resulting upon hydrolysis of pentahalides has a variable H_2O content. The salts of the acid, the antimonates, are of the type $M^ISb(OH)_6$, as Pauling showed to be necessary to conform to accepted ionic radius ratios. Although the strength of antimonic acid has not been accurately determined, it appears to be comparable to acetic acid.

Antimony(IV) oxide: Obtained by heating in air the trioxide or the hydrated pentoxide.

There is a marked structural difference between the phosphates and the antimonates. Thus sodium pyroantimonate, $Na_2H_2Sb_2O_7 \cdot 5H_2O$ contains the ion $Sb(OH)_6^-$ rather than $Sb_2O_7^{4-}$, and the magnesium compound (hydated) which has a 12:1 ratio of oxygen to antimony, and would thus be a hexahydroxyantimonate, has the (X-ray determined) structure $[Mg(H_2O)_6][Sb(OH)_6]_2$.

Sulfides: Sb_2S_3 and Sb_2S_5, which may be obtained from the elements or by precipitation, respectively, of Sb(III) and Sb(V) solutions with H_2S. The Sb_2S_3 dissolves in alkaline solutions to form thioantimonites, containing the ion SbS_3^{3-} , or $Sb(SH)_6^-$, while Sb_2S_5 forms the thioantimonates, containing SbS_4^{3-}. The latter is probably present as $[SbS_2(SH)_{21}(OH)_2]^{3-}$ or $[SbS_4(H_2O)_2]^{3-}$.

In alloys, antimony is easily detected by its formation of a white solid upon treatment with concentrated HNO_3 and subsequent separation from tin, which is the only other metal thus forming a white solid.

Both trivalent and pentavalent antimony form several organoantimony compounds. Some of these include methylstibine CH_3SbH_2 and the substitution product, methyldichlorostibine CH_3SbCl_2; phenylstibine $C_6H_5SbH_2$ and the substitution product, phenyldichlorostibine $C_6H_5SbCl_2$; methylantimony tetrachloride CH_3SbCl_4; phenylantimony tetrachloride $C_6H_5SbCl_4$; sodium methylantimonate $Na[CH_3Sb(OH)_5]$; sodium trifluoromethyl antimonate $Na[(CF_3)_3Sb(OH)_3]$; triethylstibine sulfide $(C_2H_5)_3SbS$; tetraphenylstibonium tetraphenylborate $[(C_6H_5)_4Sb][B(C_6H_5)_4]$; stibiobenzene $C_6H_5Sb{=}SbC_6H_5$ and lithium hexaphenylantimonate $LiSb(C_6H_5)_6$.

Uses: Representative alloys containing antimony are described in the accompanying table.

ANTIMONY CONTENT OF REPRESENTATIVE ANTIMONY-CONTAINING ALLOYS

Hard lead	Up to 12% Sb
Antimony reduces mp of Pb and hardens resulting alloy. Alloy has better abrasion resistance than chemical Pb at temperatures below 140°C. Alloy is age-hardenable.	
Tin-lead solders	Up to 1% Sb
Type metals	3–19% Sb
These Pb-base alloys also contain from 3-9% Sn.	
Lead-base diecasting alloys:	
[a]ASTM No. 4	14–16% Sb
ASTM No. 5	9.25–10.75% Sb
Bearing alloy	15% Sb
CT metal	12.5% Sb
Tin-free alloy	10% Sb
Babbitt (bearing) metals:	
[b]SAE 10	4–5% Sb
SAE 11	6–7.5% Sb
SAE 12	7–8.5% Sb
SAE 13	9.25–10.25% Sb
SAE 14	14–16% Sb
SAE 15	14.5–16% Sb
Britannia metal	5% Sb
This alloy also contains 93% Sn and 2% Cu. Very useful for spinning utensils.	
Pewter	Up to 7% Sb
Pewter also contains up to 20% Pb and 4% Cu with the remainder made up by Sn.	

[a] American Society for Testing and Materials

[b] Society of Automative Engineers

Metallic antimony is an effective pearlitizing agent for producing pearlitic cast iron. The principal use of antimony, however, is in the form of the oxide. Its major application is as a flame retardant for plastics and textiles. Other applications of importance are in glass, pigments, and catalysts.

Toxicity: The threshold limit value of antimony and its compounds is 0.5 milligram/cubic meter (as Sb). Antimony and its compounds used under conditions giving rise to dust, fume, and vapor should be carried out under proper ventilation. In handling antimony and its compounds, appropriate hygienic practices and good housekeeping should be observed. Stibine, SbH_3, requires extreme caution in handling because it is very toxic. When using antimony and its compounds, reducing conditions, which may give rise to the undesired formation of stibine, must be avoided.

Additional Reading

Carapella, S. C., Jr.: "Properties of Pure Antimony," in "Metals Handbook," Vol. 2, ASM International, Materials Park, Ohio. (Revised periodically.)

Perry, R. H., and D. Green: "Perry's Chemical Engineers' Handbook," 6th Edition, McGraw-Hill, New York, 1988.

Sneed, M. C., and R. C. Brasted: "Comprehensive Inorganic Chemistry," Vol. 5, Van Nostrand Reinhold, New York, 1956.

Staff: "Handbook of Chemistry and Physics," 73rd Edition, CRC Press, Boca Raton, Florida, 1992–1993.

ANTINODES (or Loops). The points, lines, or surfaces in a standing wave system where some characteristic of the wave field has maximum amplitude.

ANTIOXIDANT. Usually an organic compound added to various types of materials, such as rubber, natural fats and oils, food products, gasoline, and lubricating oils, for the purposes of retarding oxidation and associated deterioration, rancidity, gum formation, reduction in shelf life, etc.

Rubber antioxidants are commonly of an aromatic amine type, such as dibeta-naphthyl-para-phenylenediamine and phenyl-beta-naphthylamine. Usually, only a small fraction of a percent affords adequate protection. Some antioxidants are substitute phenolic compounds (butylated hydroxyanisole, di-tert-butyl-para-cresol, and propyl gallate).

When used in foods, antioxidants are highly regulated to extremely small percentages in most countries—down to the low fractions of one percent. Composition of the substrate, processing conditions, impurities, and desired shelf life are among the most important factors in selecting the best antioxidant system for a given food product. The desirable features of antioxidants may be summarized as (1) effectiveness at low concentrations; (2) compatibility with the substrate; (3) nontoxic to consumers; (4) stability in terms of conditions encountered in processing and storage, including temperature, radiation, pH, etc.; (5) nonvolatility and nonextractability under the conditions of use; (6) ease and safety in handling; (7) freedom from off-flavors, off-odors, and off-colors that might be imparted to the food products; and (8) cost effectiveness.

Mechanism of Oxidative Degradation. It could appear that inasmuch as oxidative degradation occurs in a variety of organic materials that are dissimilar in appearance and have entirely different applications and different properties, with degradation producing different effects, the oxidation mechanism itself might be different. Current knowledge indicates, however, that the mechanism of oxidative degradation is the same for all organic substances. They appear to degrade by the same free-radical mechanism.

Common examples of food oxidative degradation include products that contain oils and fats. For example, some antioxidants have made it possible to store groundnuts (peanuts) and other nuts, maize (corn) products, and bakery and cereal products on the shelf for periods well in excess of the four months that was considered the traditional limiting period prior to the appearance of such additives. Other examples of food products that tend to become rancid by way of oxidation include various meat-flavor stuffing mixes, cake mixes, unbaked cheesecake mix, and essentially all foods that incorporate lipids. The stability of natural fats and oils present in raw materials varies over a wide range and hence the amount of antioxidant required must be tailored to each product situation. Enzymatic "browning" is another example of oxidative degradation. The enzymes in fruits and vegetables cause apples, apricots, and potatoes, among others, to darken when they are exposed to air after being cut, bruised, or allowed to overmature. Some antioxidants can prevent or delay enzymatic browning much in the same manner as dipping freshly cut fruits in lemon, orange, or pineapple juice. Limonene and ascorbic acid naturally present in these juices serve as antioxidants. Oxidative changes may affect carbohydrate, protein, and fat substances, the primary building blocks of foodstuffs, but generally the oxidative rancidity problem results mainly from the *autoxidative* degradation of fatty (glyceridic) components.

Some authorities describe oxidation as a free-radical, chain-type reaction. At usual processing temperatures and more slowly at room temperature, organic free radicals (R •) are formed. These react with oxygen to form peroxy radicals (ROO •), which can abstract a hydrogen atom from the affected substance to form a hydroperoxide (ROOH) and another organic free radical. The cycle repeats itself with the addition of oxygen to the new free radical. The unstable hydroperoxides left along with the substance are the major source of degradation. Under the influence of heat, light, and any metals if present, the hydroperoxides decompose to form carbonyl groups. When this happens, the organic molecule breaks and splits off another organic free radical. Ultimately, this type of degradation can lead to rancidity and color deterioration in oils and fats.

An antioxidant ties up the peroxy radicals so that they are incapable of propagating the reaction chain or to decompose the hydroperoxides in such a manner that carbonyl groups and additional free radicals are not formed. The former, which are called *chain-breaking antioxidants, free-radical scavengers,* or *inhibitors*, are usually hindered phenols or amines. The latter, called *peroxide decomposers*, are generally sulfur compounds or organophosphites. A number of antioxidants useful in rubber and plastics, for example, are not suited to food products because of their toxicity.

A mixture of two antioxidants often will display synergism. Probably the most generally effective mixtures of antioxidants are those in which one compound functions as a decomposer of peroxides (sulfides, thiodiproprionate) and the other as an inhibitor of free radicals (hindered phenols, amines). Although the latter retards the formation of reaction chains, some hydroperoxide is nevertheless formed. If this hydroperoxide then reacts with a decomposer of peroxides, instead of decomposing into free radicals, the two antioxidants act together to complement each other. Moreover, the peroxide decomposer may itself be subject to oxidation by peroxy radicals, and its efficiency will therefore be increased in the presence of an inhibitor of free radicals. In the case of phenolsulfide mixtures, the sulfide (peroxide decomposer) also continuously regenerates the phenol (radical scavenger) to accentuate the synergistic nature of the mixture. Metal chelators or deactivators, such as citric and phosphoric acids, of prooxidant metals (iron, copper, nickel, tin), ultraviolet-light absorbers (carbon black, substitute benzophenones, benzotriazoles, and salicylates), and antiozonants (substituted phenylenediamines) also develop synergistic effects with antioxidants.

Applications of Antioxidants. The use of antioxidants in foods, pharmaceuticals, and animal feeds (direct feed additives), as well as their use in food-contact surfaces (indirect additives) is closely regulated by the governments of several countries. Antioxidants are approved only after extensive extraction, toxicological, and feeding studies. The list is relatively limited. Although antioxidants have been used for several decades and some occur naturally in food substances, intensive research in continuing, partly accelerated by the growing use of unsaturated oils in numerous food products.

Butylated hydroxyanisole (BHA) was first used in food products in 1940. This continues as one of the commonly used antioxidants, sometimes in combination with butylated hydroxytoluene (BHT), propyl gallate, citric, or phosphoric acids, to obtain a synergistic effect. In foodcontact surfaces, BHT has been used by itself or in combination with thiodipropionates and/or phosphoric acids, to obtain a synergistic effect. Well over $50 million of antioxidants are produced per year commercially in the United States alone.

The value of antioxidant protection by way of natural food sources has been pointed out in the literature with considerable frequency. Among the components of soy flour known to have some antioxidant properties are isoflavones and phospholipids. Amino acids and peptides in soybean flour also possess some antioxidant activity. There also may be some antioxidant impact from aromatic amines and sulfhydryl compounds.

Rosemary and sage have been shown to have effective antioxidant properties. The extracts in the past have been of strong odor and bitter taste and thus unsuited for use in most food products. However, solvent extraction procedures have been developed to produce purified antioxidants from rosemary and sage.

For many years, in connection with certain food products, a barrier to freeze-drying has been the problems associated with the storage stability of foods that are susceptible to lipid oxidation. In order for such foods to have a reasonable shelf life and acceptable flavor characteristics, protective additives which retard oxidation, are often added before dehydration. Such antioxidants must carry through the process and not be lost because of volatilization. For these applications, BHA, BHT, and tert-butylhydroquinone (TBHQ) have been found quite effective.

Additional Reading

Bigelow, S. W.: "Food Chemicals Codex: A Progress Report," *Food Technology*, 88 (May 1991).

Burdock, G. A. et al: "GRAS Substances," *Food Technology*, 78 (February 1990).

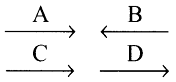
Dougherty, M.: "Synthetic Antioxidants," *Ingredient Technology IFT Short Course*, Atlanta, Georgia, March 1, 1991.

Evans, R. J.: "Alternatives to Synthetic Antioxidants," *Ingredient Technology IFT Short Course*, Atlanta, Georgia, March 1, 1991.

Staff: "Sulfites in Foods," *Food Technology*, 48 (September 1986).

Staff: "Antioxidants," *Food Technology*, 94 (September 1986).

ANTIPARALLEL. Having opposite senses. Thus the vectors **A** and **B** are antiparallel, while the vectors **C** and **D** are parallel.



ANTIPARTICLES. One of the great discoveries of modern physics is that for every type of elementary entity of matter and radiation (particle), there exists a corresponding conjugate type of entity (antiparticle). In the antiparticle, certain of the particle-defining properties are identical (*conjugation-invariant*) and others are reversed in sign (*conjugation-reversing*). The reversed sign in a conjugation-reversing property allows one to maintain a conservation law for that property in the dramatic processes of *pair creation* and *pair annihilation* in which an antiparticle is observed to appear and disappear together with the particle to which it is conjugate. In those cases where all the conjugation-reversing properties occur with zero values, the antiparticle is identical with the particle. The progressive recognition of the existence of antiparticles was initiated by Dirac's relativistic antielectron theory in 1931, and by Anderson's independent experimental discovery of the antielectron (the positron) in 1932. See also **Particles (Subatomic).**

ANTIPODAL CELLS. The three usually small cells which occur in the embryo sac of angiosperms at the end most distant from the micropyle. No known function has been ascribed to them.

ANTIPODE. With respect to any given point, the opposite point is known as the antipode. In terms of a sphere, Y would be the antipode of X of a line XY drawn through the center of the sphere and intersecting the two opposite surfaces. Antipodes often refer to regions rather than specific localized areas. Thus, the British Isles are approximately antipodal with Australia and New Zealand.

ANTIPROTON. An elementary particle having a mass equal to that of the proton, differing from the proton only in the sign of its charge, which is negative, and a magnetic moment oppositely directed with respect to its spin. Positive identification of the antiproton was first made at the University of California. In 1959 Segre and Chamberlain received the Nobel Prize in physics for this discovery. Protons which had been accelerated to an energy of 6200 MeV in the bevatron, the proton synchrotron of the University of California Radiation Laboratory at Berkeley, were allowed to collide with a copper target. Negatively charged particles coming out in a forward direction from this collision were selected and separated in momentum by a focusing and analyzing magnet system to provide a beam of negative particles of known momentum. After a time of flight of about one-tenth of a microsecond, this beam may be expected to consist mainly of negative pions and muons, with some negative kaons (mass about 965 electron masses) and possibly negative protons. These particles were then distinguished both by measurement of their time of flight from the target (since particles of different mass have different velocities for given momentum) and by means of a device measuring the velocity of each particle passing through by the angle of its Cerenkov radiation. In this way the presence of negative particles with protonic mass (within about 10%) and distinct from the known kaons and hyperons was established. Their rate of production for the momentum and direction of this experiment was about one negative proton for every 50,000 negative pions with the same momentum and direction.

Antiprotons were first captured (1987) by researchers working with the Low-Energy Antiproton Ring (LEAR), which is a part of the large European Laboratory for Particle Physics (CERN) located in Geneva, Switzerland. The team of scientists made up from physicists of the University of Washington and the University of Mainz (West Germany) captured antiprotons from a high-energy accelerator and stored them for several minutes in an electromagnetic ion trap. The scientists forecast that with improvements in the trapping process it will facilitate the precise measurement of the inertial mass of the antiproton. They plan to ascertain the mass from the frequency of the circular motion of the antiprotons around the magnetic field lines in the trap (cyclotron resonance frequency). This will require slowing down the axial motion in the trap from the kiloelectron volt energies of the present equipment to a maximum of approximately 5×10^{-4} eV, i.e., the thermal energy associated with the ambient temperature of 4.2 K. To make the most meaningful measurement, there should be just one antiproton cooled to 4.2 K and the trapping time should be at least one day, compared with the minutes of the first experiment. Much greater vacuum and cryogenic cooling with helium is planned for future experiments. The scientists envision a number of interesting uses for trapped antiprotons—determination of the gravitational constant of an antiproton; possible use of large numbers of antiprotons as an energy source for space and military applications, although the scientists at this juncture (mid-1987) admit that collecting antiprotons in macroscopic quantities stretches one's imagination. See also **Particles (Subatomic);** and **Proton.**

ANTIPYRETIC. Any physical agent or drug that lowers the temperature of the body. Among antipyretics used are aspirin, antipyrine, acetanalid, and phenacetin. See also **Analgesics.**

ANTIRESONANCE (or Parallel Impedance). In general, a condition of maximum impedance, as results when two or more impeders are connected in parallel, under such conditions that (for resistanceless impeders), Z approaches infinity when $\omega_1^2 = 1/LC$.

ANTISOLAR POINT. The point on the celestial sphere that lies directly opposite the sun from the observer, i.e., on the line from the sun through the observer. See also **Celestial Sphere and Astronomical Triangle.**

ANTISTOKES LINES. When all the molecules are in the normal state, the transitions caused by exciting radiation are those in which the scattered or fluorescent light is of the same or lower frequency than the incident light (Stokes Lines). However if some of the atoms or molecules are in states other than normal, lines of frequencies higher than that of the incident light (Antistokes Lines) may result. See **Raman Spectrometry.**

ANTISYMMETRIC. In mathematics, a term used to denote a function which is transformed into its negative when the variables of the function are interchanged in pairs. In physical science, the term antisymmetric is applied to any physical system in which each point has properties opposite to those of a point symmetrically located with respect to it, e.g., an electric dipole is antisymmetric in its charge distribution. See **Symmetric.**

ANTITOXIN. (1) A substance made and elaborated in the body to neutralize a specific bacterial, plant, or animal toxin; (2) one of the class of specific antibodies. See also **Antibody.**

The history of the development of antitoxins in combating bacterial infection dates back to the early beginnings of organized bacteriology. Behring was the first to show that animals that were immune to diphtheria contained, in their serum, factors which were capable of neutralizing the poisonous effect of the toxins derived from the diphtheria bacillus. While this work was carried out in 1890, prior to many of the great discoveries of mass immunization, and much later the antibiotics, there yet remains a place for antitoxins in medical treatment or prophylaxis for some diseases, such as tetanus and botulism.

The more important approach to immunity to infectious disease now is the development of *active immunity* by injection of a vaccine. The vaccine may be either an attenuated live infectious agent, or an inactivated or killed product. In either case, protective substances are generated in the bloodstream called antibodies which help to neutralize the infectious agent when it is introduced. The principle of *passive immunization*, on the other hand, involves the development of the antibodies in another host and most frequently a different species as well. The antiserum or antitoxin (from the other host) is employed in preventing the onset of the disease, or in actual treatment of the active infection in subjects who have not had the advantage of becoming actively immunized due either to neglect or to the fact that an effective vaccine was not available. The use of antitoxins prepared in another species (for example, horse) is not without some element of risk.

Antitoxins are prepared by injecting the donor animals with frequent and increasing doses of toxin while maintaining a level at each injection that the animal can tolerate. The initial doses are critical since these toxins may be among the most poisonous agents known. In one technique, the toxin is diluted so that the first injection contains less than the minimum lethal dose. Other programs used toxins that are inactivated with formaldehyde so that they are no longer poisonous, but may still elicit an immune response and result in antitoxin that will neutralize the unaltered toxin. This method of inactivation also was developed in the nineteenth century for preparing many of the important vaccines against diseases, such as influenza, tetanus, and diphtheria.

Small laboratory animals are helpful in carrying out research in this field, but in commercial production, larger animals are required. The horse proved to be very satisfactory for large-scale production of antitoxins, and is still used for the major portion of antitoxin production, although some material is also derived from cattle. The management of the animals and the schedules and dosages are perhaps as much an art as an exact science. The selection of the horses that have the best potential as antitoxin producers is also a critical factor. When a horse is receiving the maximum level of toxin in the hyperimmunization program, it may be injected in a single dose with enough toxin to be fatal for 100 million mice if the material was suitably distributed.

A means for enhancing the potency of the antitoxin in horses is to use an agent called an *adjuvant*. Several adjuvants have been used, including tapioca, mineral oil, and aluminum hydroxide. The mechanism by which these agents increase the intensity of the immune response is not fully understood, but local inflammatory reaction and the resulting slower release of the injected material from the original site appear to play a role. Adjuvants also are sometimes used with vaccines for human use. It has been recorded that one horse, during an eleven-year period, gave 657 gallons ($\sim$25 hectoliters) of blood from which tetanus antitoxin and, at different times, pneumococcus antiserum, was prepared. The volume of serum removed from the horse can be increased by the return of the red cells after removal of the plasma or liquid component. It has been shown that if the red cell level can be maintained in human donors of special sera, they can safely give as much as one liter of serum per week.

The injection of serum components of another species into human patients has not been without problems. A condition known as serum sickness develops in an alarmingly high proportion of those treated in this way, which is apparently due to a generalized sensitivity which develops to the foreign serum protein. The onset of illness is usually delayed for several days after the injection of antitoxin and symptoms may be quite severe.

See also **Immune System and Immunology.**

A. C. Vickery, Ph.D., Associate Professor, College of Public Health, University of South Florida, Tampa, Florida.

ANTIVIRAL DRUGS. Compounds for use in the treatment of viral infections are quite limited. A problem in the development of antiviral drugs centers around the relationship between the replication function of the virus and functions of host cells. Obviously, a drug must target the virus and virus-infected cells with no destruction of healthy cell functions. Among antiviral drugs currently used are amantadine hydrochloride, idoxuridine, and adenine arabinoside. Research activity in the antiviral drug field is vigorous because of the obvious great need for them.

Amantadine Hydrochloride. Chemically, this drug is 1-adamantanamine hydrochloride. A commercial preparation is known as *Symmetrel®*. This drug is indicated in the preventing (prophylaxis) and symptomatic management of respiratory tract illness caused by influenza A virus strains. The drug is particularly considered in connection with high-risk patients, close household or hospital ward contacts of index cases, and patients with severe influenza A virus illness. In the prophylaxis of influenza due to A virus strains, early immunization as periodically recommended by public health authorities is the method of choice. When early immunization is not feasible, or when the vaccine is contraindicated or not available, amantadine hydrochloride is sometimes used chemoprophylactically with inactivated influenza A virus vaccine until protective antibody responses develop. Principal contraindications include hypersensitivity to the drug and patients with a history of epilepsy, congestive heart failure, or peripheral edema.

To date, amantadine hydrochloride has not been used extensively clinically, although excellent controlled trials indicating that it has a prophylactic effect have been reported. Infected persons have been reported to suffer less cough, sore throat, and fever than other persons not taking the drug. Reports also indicate the drug accelerates recovery from peripheral airway abnormalities in normal persons who have uncomplicated influenza.

In the treatment of mild parkinsonism, amantadine has been used to advantage, particularly in conjunction with L-dopa. It is believed that amantadine stimulates the release of dopamine from nerve terminals.

Side-effects from the continuous use of amantadine include dizziness, drowsiness, difficulty in thinking, hallucinations, convulsions, and, rarely, psychosis. The effects are more commonly experienced by middle-age groups and the elderly.

Idoxuridine. Chemically, this drug is 5-iodo-2′-deoxyuridine. This drug is indicated for the treatment of herpes simplex keratitis. The drug inhibits replication of herpes simplex virus by irreversibly inhibiting the incorporation of thymidine into the viral DNA. Although tests with rabbits of known genetic ancestry have been completed with no malformations resulting from idoxuridine, the drug is still administered with caution in pregnancy or in women of childbearing potential. The commercial compound is available in the form of a solution or as an ophthalmic ointment, with petrolatum used as an inactive ingredient. While the compound will frequently control infection, it apparently has no effect on the accumulated scarring, vascularization, or on the resultant progressive loss of vision.

Adenine Arabinoside. Also known as vidarabine, or Vira-A®, has the empirical formula, $C_{10}H_{13}N_5O_4 \cdot H_2O$. The chemical name is 9-β-D-arabinofuranosyladenine monohydrate. This compound is indicated in the treatment of herpes simplex virus encephalitis. Controlled studies have indicated that therapy with this drug may reduce the mortality caused by herpes simplex virus encephalitis from 70 to 28%. The therapy does not appear to alter morbidity and resulting serious neurological sequelae in the comatose patient. Thus, early diagnosis and treatment are essential. Herpes simplex virus encephalitis should be suspected in patients with a history of an acute febrile encephalopathy associated with disordered mentation, altered level of consciousness, and focal cerebral signs. Licensing for this use (United States) was made in December 1978. The drug was licensed in 1977 for treatment of acute herpes simplex keratoconjunctivitis and recurrent epithelial keratitis caused by herpes simplex types 1 and 2. Experimentally, under controlled conditions, the drug has been used for treatment of herpes simplex type 1 encephalitis, herpes simplex type 2 infections, herpes zoster, smallpox, progressive multifocal encephalopathy, and chronic hepatitis B virus and cytomegalovirus infections. Large doses of adenine arabinoside administered parenterally have been noted to suppress bone marrow function, particularly if administered over long periods.

Acyclovir. Some authorities consider this the most potent antiviral metabolic antagonist. Intravenous acyclovir was approved in November 1982 for immunocompromised patients with mucocutaneous herpes simplex infection and for immunocompetent patients with initial genital herpes infection severe enough to require hospitalization. Acylclovir is disease specific and is reported to have low toxicity, mainly because the activity of the drug depends on the presence of a deoxycytidine kinase that is synthesized only in cells infected by herpes simplex or varicella zoster. This enzyme, rather than the host thymidine kinase, catalyzes metabolism of the drug to the actual inhibitor, the triphosphate derivative of acyclovir. Kidney function is monitored closely when the drug is used, particularly in patients who have preexisting renal disease.

Still considered experimental among antiviral drugs and techniques are methisazone, human leukocyte interferon, smallpox vaccination and the photoinactivation of virus with light-sensitive dye. These latter techniques are considered potentially dangerous as well as being ineffective in the treatment of herpetic disease.

ANTLERITE. Antlerite is a relatively uncommon mineral found within the oxidized zones of copper deposits in arid regions. It is a basic sulfate of copper, $Cu_3(SO_4)(OH)_4$, crystallizing in the orthorhombic system. Hardness of 3.5, specific gravity 3.88, with vitreous luster, and emerald-green to black-green color.

Originally found in Arizona. It is the principal copper ore mineral at Chuquicamata, Chile.

ANT LION (*Insecta, Neuroptera*). *Immature insects of the family Myrmeleonidae* which lie buried at the apex of conical pits in dry sand or dust. Ants or other small insects which enter the pit slide down the loose slope and are seized by the upturned jaws below.

These insects are black or brown in color. They develop within a cocoon made of silk threads developed from larvae at the insect's tail. The immature insect remains in the cocoon stage until it becomes an adult.

ANT-LOVING CRICKET (*Insecta, Orthoptera*). *Mymecophilia*. Small peculiarly formed crickets which live in ant nests.

ANTONOFF RULE. The tension at the interface between two saturated liquid layers which are in equilibrium is equal to the difference between the individual surface tensions against air or vapor of the two saturated solutions. This rule is approximate only and a number of exceptions are known.

ANURA. The frogs, toads, and allied species; a division of the *Amphibia* characterized by the absence of the tail and a tadpole larval stage. Also known as the Salientia from their jumping powers.

ANUS. The terminal portion of the rectum, which forms an external opening. See also **Digestive System (Human).**

AORTA. The main and largest blood vessel of the arterial blood system. It arises from the left ventricle of the heart, and arching over the root of the left lung, descends along the vertebral column, passing through the chest, and pierces the diaphragm into the abdominal cavity, finally dividing into the right and left iliac arteries in the pelvis. The branches of the aorta supply oxygenated blood to every part of the body. See **Arteries and Veins (Vascular System);** and **Heart and Circulatory System (Human).**

APATITE. The mineral apatite is a phosphate of calcium with either fluorine or chlorine or sometimes both, hence the distinction between fluorapatite and chlorapatite. Sometimes both fluorine and chlorine are present. Most apatite is, however, fluorapatite, $Ca_5(PO_4)_3F$.

Apatite crystallizes in the hexagonal system in prismatic and tabular forms. Hardness, 4.5–5; specific gravity, 3.17–3.23; luster, vitreous to resinous; transparent to opaque; streak, white; cleavage, imperfect basal and prismatic; color, white, green, yellow, red, brown and purple; subconchoidal fracture. The variety called asparagus stone is yellow-green and manganapatite, which is a dark bluish-green, may contain as much as 10% manganese dioxide replacing the calcium. Werner devised the name *apatite* from the Greek word meaning "to deceive", because it was frequently mistaken for beryl and other species. Apatite has been found widely distributed both geographically and petrologically as it occurs in many sorts of rocks, metamorphic limestones, gneisses, schists, granites and syenites, pegmatite veins and even with iron ores. It has been prepared artificially. It has been mined for the manufacture of fertilizers and to a slight extent for jewelry.

Apatite occurs extensively in Europe and America, especially in New England, New Jersey, New York, North Carolina, California, and in the provinces of Ontario and Quebec in Canada.

APERTOMETER. A device designed by Abbe for measuring the numerical aperture of microscope objectives.

APERTURE. Qualitatively, any opening through which radiation or particle fluxes may pass. In optical instruments, in communications and in other fields, the term aperture has acquired various specific meanings. For example, the aperture of a lens is simply the diameter of the lens. However, the numerical aperture (N.A.) of a lens is the quantity $n \sin u$, where u is the angular radius of the lens as seen from a point on the optical axis at the object, and n is the refractive index of the medium between the object and the lens. The relative aperture (or f-number) is the ratio of the focal length of an optical system to the diameter of the entrance pupil. A related term is the aperture angle, which is the angle subtended by the radius of the entrance pupil of an optical instrument at an axial object point.

The aperture of a unidirection antenna is that portion of a plane surface near the antenna, perpendicular to the direction of maximum radiation, through which the major part of the radiation passes.

APERTURE CARD. A punched card (see **Input/Output Devices)** which may contain information solely for machine storage and retrieval, this being the normal interpretation of the term. In addition, however, an aperture card may contain information that is not entered into electronic storage, such as small reproductions of diagrams, photos, etc. Thus, some of the information on the card may be fully electronically processable, whereas other information requires manual scanning and interpretation.

APES. See **Anthropoids.**

APHANITE. An aphanite is any fine-grained igneous rock whose constituents cannot be distinguished with the naked eye. The term is derived from the Greek, meaning invisible. The adjective aphanitic is applied to these rocks as well as their fine-grained groundmasses.

APHASIA. Defect of the language function due to brain damage, usually manifested in all four language modalities—speech production, speech comprehension, reading, and writing. Aphasia must be distinguished from disturbances of voice production, such as dysarthria, from poverty of speech due to intellectual impairment, from language abnormalities as in schizophrenia, and in hysterical mutism.

More than 90% of normal right-handed people have language function represented in the left cerebral hemisphere and damage of that will render them aphasic. Since normal left-handers have 70% of language function in the left hemisphere and 20% in the right, damage to either side can lead to varying degrees of aphasia.

Within the left hemisphere, the most important areas for language are Broca's and Wernicke's areas and these are the most commonly affected in strokes or various forms of cerebral thrombosis or direct brain trauma.

An excellent and detailed review of aphasia is given by Antonio R. Damasio, M.D., in the February 20, 1992, issue of the *New England Journal of Medicine*, pages 531–539.

R.C.V.

APHELION. The point in the orbit of a planet or other member of the solar system, except a satellite, where the object is most remote from the sun. It is the point on the line of apsides (see **Orbit (Astronomy))** diametrically opposite to periphelion.

APHID OR PLANT LOUSE (*Insecta, Homoptera*). A small delicate insect with sucking mouth. The aphid generally lives on the sap of plants and is of major economic importance. The aphid is characterized by an intricate life cycle that results in a high rate of reproduction. In the temperate zone, aphids hatch in the spring from winter eggs; these individuals are females known as stem mothers. They bear living young without mating (viviparity, parthenogenesis) and these, in turn, are females capable of the same type of reproduction. Late in the season, a generation known as the *sexuparae* bears both male and female offspring, which mate to produce the eggs that pass the winter.

Winged aphids (Fig. 1) appear under conditions that demand migration from plant to plant. Experiments have shown that the appearance of wings is a response to definite environmental conditions, complex in nature, and not fully understood. The much more common wingless aphid is shown in Fig. 2.

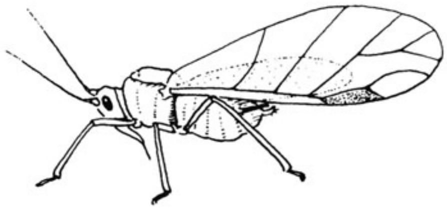


Fig. 1. Winged aphid.

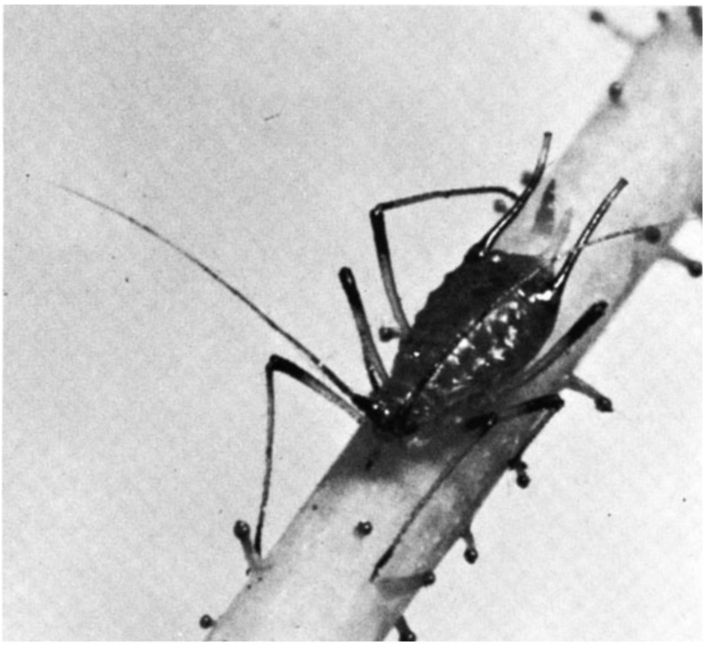


Fig. 2. Wingless aphid feeding on rosebush. (*Bernard L. Gluck from National Audubon Society.*)

The aphid injects a poisonous saliva into plants which causes discolored and unhealthy foliage and buds. A by-product of the aphid's digestion is known as honeydew, a sweet secretion arising from the ingestion of excessive sap and sugars. Ants, bees, and wasps are vigorously attracted by this honeydew. To encourage production of larger quantities of the substance, an ant has been observed to assist an aphid by carrying the aphid in its mouth from a withered plant to a healthy plant, thus encouraging the aphid to eat more and produce more honeydews. The ant also protects the aphid from intruders. The honeydew clogs the pores of plants and also nourishes damaging fungi.

Although the aphid is destructive to many crops, it is particularly damaging to certain vegetables, including bean, cabbage, cucumber, melon, pumpkin and squash; and to certain fruit and nut crops, including apple, apricot, peach, pear, pecan, and walnut. The aphid is quite damaging to a number of decorative plants.

There are many species of aphid, nearly a species for each crop attacked, indicating the high specialization created over the centuries. Generally, all aphids are tiny ($\frac{1}{16}$-inch; 1.5 millimeters) in length, ranging in color from light-green to dark-green and black. Most varieties have a rather prominent rearward protuberance, often called a "tailpipe." Although the damage varies from one crop to the next, generally the aphid causes leaves to curl and thicken, turn yellow, and die.

Most aphids overwinter in the egg stage; eggs are laid on the twigs and bark of the trees. To kill these eggs and prevent a rapid buildup of aphids later, it is well to use a dormant oil spray. Oil sprays should be applied before any color is shown in buds of plant. Dormant oil sprays must not be used when foliage is on the tree or plant, or when the temperature is higher than 85°F (29.4°C) or lower than 35°F (2°C). When aphids become numerous during the growing season, sometimes they can be washed off with soap formulations applied under fairly high pressure, but specific treatment may involve stronger chemicals, the selection varying from one infested plant to the next.

To control root-feeding aphids, such as the *woolly apple aphid*, the soil immediately under the plant should be regularly loosened and cultivated to a depth of 1 to 3 inches (2.5 to 7.5 centimeters). Attacks by the woolly aphid can be avoided by planting varieties that are resistant. Also, trees and plants that have been kept in a vigorous growing condition are less likely to suffer damage from these insects. Woolly apple aphids also feed in wounds of trunks and branches. The wounds should be treated with a wound dressing to prevent attack by aphids.

When natural enemies of the aphids are abundant, aphids usually do not require additional control measures. Before applying chemicals, check the plants carefully to see if a natural control agent, such as the larvae of lady beetles, is present. Syrphid flies and ant lions also are natural enemies of the aphid. When the spring is warm, natural enemies are usually present in sufficient numbers to control aphids completely. The greatest damage occurs after a cold spring; aphids increase more rapidly than their natural enemies in a chilly season.

Indicative of the highly specialized aphids are:

Bareberry plant louse (*Rhopalosiphium berberidis*)
Beet aphis (*Pemphigus betae*)
Birch aphis (*Callipterusbetulaecoleus*)
Black aphis (*Aphis persicae-niger*); affects peach
Bud aphis (*Siphocoryne arenae*); affects apple
Cabbage aphis (*Aphis brassicae*)
Corn-root aphis (*Aphis maidiradicis*) See entry on **Ant.**
Currant aphis (*Myzus ribis*)
Gall aphis (*Phylloxera sp.*); affects spruce
Leaf or green aphis (*Aphis pomi*); affects apple
Lettuce aphis or green fly (*Macrosiphum lactucae*)
Lupine aphis (*Macrosiphum albifrons*)
Melon aphis (*Aphis gossypii*); also affects squash, watermelon
Oleander aphis (*Aphis nerii*)
Parsnip louse (*Hyadaphis pastinacae*)
Pea aphis (*Macrosiphum pisi*); also affects bamboo
Root louse (*Aphis forbesii*); affects strawberry
Rosy aphis (*Aphis sorbi*); affects apple and quince
Spinach aphis (*Myzua persicae*); also affects green peaches
Squash aphis (*Nectarophora cucurbitae*)
Woolly aphis (*Schizoneura lanigera*); affects apple.

Oriental black citrus aphid (Toxoptera citricada). This insect is a serious pest of citrus in South America. The oriental black citrus aphid is black and, when fully grown, is about $\frac{1}{16}$-inch (1.5 millimeters) long. The antennae and legs are sometimes brownish, rather than black. The aphids generally are wingless. Ocassionally, winged forms appear. The winged aphids have transparent wings; the forewings are much longer than the rear wings. The aphids are most numerous in spring and early summer. Development takes about 12 days. These insects feed on leaves of citrus and cause new growth to be stunted. A heavy infestation can cause the trees to become seriously damaged. Besides damaging the trees directly, these aphids carry *tristeza*, a citrus disease that kills the roots and causes the trees to die.

APHIS-LION (*Insecta, Neuroptera*). The larva of the golden-eyes or lacewing flies (family *Chrysopidae*), so called because they feed on aphids and other small insects.

APHYTIC. A paleobotanic division of geologic time, signifying the period of time that preceded the first occurrence of plant life. See also **Paleobotany.**

APICAL GROWTH. Growth at the tip of an organ, as occurs in the roots and stems of all higher plants. Examination of the stems of most plants will show that growth in length occurs only in the apical portion, and only for a relatively short period of time, usually a matter of a few weeks. This may be determined by observing the distances between successive leaves: near the growing tip the leaves are very small and close together; as one goes back along the stem the size of the leaves increases and also the distance between them; but after the leaf is mature little elongation of the stem occurs, as shown by the uniform distances between the mature leaves. Older stems increase constantly in diameter, but only exceptionally in length. One notable such exception, known as intercalary growth, is found in grasses and some other plants. Here a group of cells in the region of the node are capable for a time of active division and of causing increase in the length of older portions of the stem. Both here and in the tip of the stem increase in length is due to elongation of cells produced by an actively dividing tissue known as a meristem.

For related topical coverage in this volume, see alphabetical index.

API GRAVITY. See **Petroleum; Specific Gravity.**

APLACOPHORA. A class of *Amphineura (Mollusca)* made up of animals of worm-like form with foot reduced to median ventral ridge and mantle enlarged.

APLANATIC LENS SYSTEM. A system which satisfies the equation, $ny \sin u = n'y' \sin u' =$ constant. Unprimed letters refer to object space, primed to image space. n, n' are indices of refraction; y, y' are distances of a point object and its point image from the optical axis; and u, u' are angles made by a ray with the optical axis. The equation is known as Abbe's sine condition.

A triple *aplanat* is a compound lens made of two diverging lenses of flint glass between which is cemented a converging lens of crown glass.

APLANATIC POINTS. Two points on the axis of an optical system which have the property that rays proceeding from one of them shall all converge to, or appear to diverge from, the other. The two foci of an ellipsoid of revolution are aplanatic points. An aplanatic surface is an optical surface for which two aplanatic points exist.

APLITE. This term is applied to fine-grained, sometimes sugary-textured igneous rocks, composed almost wholly of quartz and feldspar. Except for size of grain, aplites resemble permatites both in mineral composition and in mode of occurrence in dikes and veins, save that the rare minerals often present in pegmatites are wanting here. The word aplite is derived from the Greek word meaning simple, referring to its ordinarily simple mineral composition.

APODEME. An internal projection of the hard outer covering (exoskeleton) of arthropods. Apodemes provide muscle attachments and in some species are extensively developed. They are collectively termed the endoskeleton.

APODIFORMES (*Aves*). This order includes the swifts and hummingbirds, which comprise those birds that have the longest primaries in proportion to their total length. Perhaps both forms represent split-offs from an ancestral "swift-sparrow" stock occurring only after the rollers and their relatives (*Coraciformes*) had already separated from the original common stock of the arboricoles, which yet more primitively had encompassed the range from the ancestral owls to the ancestral *Passeriformes*.

Swifts and hummingbirds have several characteristics in common. The wings are flat and lack, or possess only very short supporting feathers (to cover the space between wings and body). They have four complete pairs of ribs; the posterior edge of the sternum is not notched. The legs are short, the leg muscles are very weak, and the talons are pointed. The eggs are relatively elongated and are generally white.

Despite their German name "sail birds," none of the species of the swifts can actually sail or soar in the air, making use of rising convection currents. Aside from the already mentioned characteristics they share with hummingbirds, swifts have the following features. The length is 10–33 centimeters (4–13 inches). The bill is short and broad, the opening is wide, and the salivary glands are large. The tongue is small and triangular. They have 7–10 secondaries; plumage ranges from gray to black and brown (in one tree swift species, it is olive and blue), and often has white markings. The iris is brown, and the bill and feet are black. The clutch numbers from 1–6, and both sexes incubate and feed.

Swifts occur in all parts of the world except the polar regions. There are two families: 1. The Swifts (*Apodidae*); and 2. the Tree Swifts (*Hemiprocnidae*); there is a total of 77 species (from 65–81, according to other authorities).

One of the most clearly characterized, most sharply defined, and most homogeneous families is the hummingbirds (*Trochilidae*). Hummingbirds are distinguished from swifts by the following characteristics. They range in size from small to very small, with a length of 6–22 centimeters ($2\frac{1}{2}$–$8\frac{1}{2}$ inches) and a weight of 2–20 grams (0.07–0.7 ounce). The large eyes are set in wide sockets. The bill is usually long, narrow, tube-shaped, straight or curved, and very thin throughout its length. Although the neck is long and flexible, the head is usually retracted between the shoulders. The wings are long to very long. The bird has 10 well-developed primaries, with the outermost almost always being the longest, but it has only 6 (rarely 7) secondaries. The origins of the powerful flight muscles are on a particularly well-developed sternum. The shins and feet are short, the leg and foot bones are thin, but the toes, of which there are 3 forward and 1 reversed, are well developed, short, with sharply pointed talons. The feet are suitable only for perching on branches, not for walking. The body plumage is moderately dense and the feathers are firm and adherent to each other, not downy.

Despite the richness of species, the family of the hummingbird is one of the most homogeneous known. The subdivision into tribes, genera, subgenera, species, and subspecies is therefore extremely difficult. There are approximately 120 genera with about 400 distinct forms, of which we consider 321 to be species. They occur only in America. See also **Swifts and Hummingbirds.**

APONEUROSIS. A sheet of tough white glistening fascia or membrane which surrounds muscle and muscle fibers or connects a muscle to the part which it moves. The tensile strength and resistance of muscle tissue are dependent upon the fascial tissue around the muscular fibers. See also **Tendon.**

APOPHYLLITE. The mineral is a hydrous silicate of potassium, calcium, and fluorine, corresponding to the formula, $KCa_4Si_8O_{20}(F,OH)\cdot 8H_2O$. The true crystallographic symmetry is evident on crystals by the luster difference between the basal pinacoid facial planes and other crystal faces. Also, prism faces show vertical striations; basal planes do not. It crystallizes in the tetragonal system in square prisms resembling cubes terminated by based pinacoid or pyramids, often with both. Prism faces show vertical striations; basal pinacoid either dull or rough. Cleavage is perfect, parallel to the base; hardness, 4.5–5; sp gr, 2.3–2.4; luster, vitreous to pearly; transparent to translucent or nearly opaque; color may be white, grayish, greenish, yellowish, or reddish. This mineral was named by Haüy from the Greek words meaning from a leaf, referring to its exfoliation when heated with the blow pipe.

Apophyllite is a secondary mineral found with the zeolites and has been classed with them by some writers, but it contains no aluminum, which element is understood to be an essential in a zeolite. It occurs in cavities in basalts and less often filling openings in granites or other crystalline rocks; it also is a gangue mineral in certain ore veins.

There are many localities for apophyllite: Bohemia, Trentino, Italy, the Hartz Mountains, and Iceland. Fine specimens have been obtained from the Ghats Mountains in India. The Triassic trap rocks of New Jersey, Connecticut, and Nova Scotia have also furnished many specimens.

APOPHYSIS. In zoology, an apophysis is a protuberance or outgrowth of an organic structure, such as a process on a bone.

In geology, an apophysis is a tongue or other direct offshoot of a larger vein or dike.

APPARENT MOLAR QUANTITY. For a solution containing n_1 moles of solvent and n_2 moles of solute, an apparent molar quantity is defined as

$$\frac{X - n_1 x_1}{n_2}$$

where X is the value of the quantity for the whole solution and x_1 the molar quantity for the pure solvent; e.g., the apparent molar volume of the solute is

$$\frac{V - n_1 v_1}{n_2}$$

V being the total volume and v_1 the volume per mole of the solvent.

See also **Molal Concentration; Molar Concentration; Mole (Stochiometry); Mole Fraction; Mole Volume.**

APPENDAGE. A supplementary structure attached to an organ or body. Externally many animals have appendages which serve for defense, for locomotion, for securing food, and as sensory organs.

The simplest external appendages are mere outgrowths of the body wall, either solid or hollow; the tentacles of coelenterates are an excellent example. In the annelid worms both tentacles and parapodia appear as external appendages, in the echinoderms the rays or arms may be radiating divisions of the body or appendages, and in the mollusks tentacles of complex structure occur. In all cases these structures allow greater facility of movement than is possible for the body as a whole and so compensate the lack of freedom which attends increased size and complexity or sessile habits.

The most elaborate appendages are the joined appendages of the arthropods and chordates. Although they are not fundamentally related, the appendages of both phyla are similar in principle. They consist of a series of segments connected by movable joints with each other and with the body, and are provided with muscles which operate them as a series of levers. In arthropods the skeleton is external to the muscles, in the vertebrates internal.

The jointed appendages of arthropods are specialized in various species for swimming, walking, running, jumping, feeding and grasping, and in the form of antennas and palpi as sensory organs. In the jumping appendage powerful muscles result in the extension of one segment of the leg, as in the familiar grasshoppers. Grasping is accomplished by the folding back of one segment against another in the insects or by the development of a process on the next to the last segment which works against the terminal segment in a forceps-like relation; the latter is the chelate type of appendage. Arthropod appendages were originally metameric, one pair appearing on each segment of the body. In existing species they are variably limited as described under the term *Arthropoda*.

The vertebrate appendage is also specialized for various purposes, including jumping, swimming, flight, and burrowing. The primitive form appears to be the paired fins of the fishes (*Osteichthyes*), and the arms and legs of humans are good examples of fairly specialized appendages. Two pairs, the pectoral or anterior pair and the pelvis or posterior pair are typical; only in highly specialized forms such as the snakes is this number reduced.

The vertebrate appendage differs from the arthropod appendage in the terminal series of digits. These structures have been developed in some animals so that they can be opposed to each other for grasping, as in the opposition of the human thumb to the four fingers. In other forms they have been reduced in importance.

See also **Biramous Appendage; Pentadactyl Appendage.**

APPENDICITIS. Inflammation of the vermiform appendix. See **Appendix.** The infection may be acute and lead to perforation of the appendix if not promptly attended, or it may subside spontaneously. Pain, tenderness, and spasm in the right lower quadrant of the abdomen are typical symptoms of appendicitis—but these are also symptoms of other conditions, such as duodenal ulcer disease, cholecystitis (inflammation of the gallbladder), gastrointestinal actinomycosis, and gastroenteritis resulting from *Salmonella* infection, among others. The symptoms of acute appendicitis develop rapidly into a severe generalized abdominal pain, and rebound tenderness is often present. In children, crying, vomiting, and refusal of food may be very early symptoms of appendicitis. Because such symptoms commonly occur from dietary indiscretions, parents may administer a laxative before consulting a physician, with the possible result that perforation of the appendix may be precipitated. Thus, a laxative or enema is not advised in the case of abdominal pain without first consulting a physician.

Immediate surgery (*appendectomy*) is indicated in acute appendicitis. Delay invites complications, such as peritonitis. See **Peritonitis.** Devoid of complications, the prognosis for successful recovery from appendicitis is excellent.

Additional Reading

Crabbe, M. M., et al.: "Recurrent and Chronic Appendicitis," *Surg Gynecol Obstet.*, **163** (1986).

Fitz, R. H.: "Perforating Inflammation of the Vermiform Appendix; with Special Reference to Its Early Diagnosis and Treatment," *Amer. J. Med. Sci.*, 321 (1886). (A classic reference.)

Lewis, F. R, et al.: "Appendicitis: A Critical Review of Diagnosis and Treatment in 1,000 Cases," *Arch. Surg.*, **84,** 110 (1975).

Scully, R. E., et al., Editors: "Case 35–1990 (Appendicitis)," *N. Eng. J. Med.*, 593 (August 30, 1990).

Scully, R. E., et al., Editors "Case 22–1991 (Appendicitis)," *N. Eng. J. Med.*, 1575 (May 30, 1991).

Taveras, J. M. and J. T. Ferrucci, Editors: *Radiology: Gastrointestinal, Abdominal and Pelvic, Genitourinary*, J. P. Lippincott, Philadelphia (1989).

APPENDICULARIAN (*Chordata, Tunicata*). A free-swimming tailed tunicate with a permanent chordate tail. These forms make up the class *Larvacea*.

APPENDIX (or Appendix Vermiformis). A blind worm-like tubular portion of the intestine which arises from the base of the cecum. Its size and position vary greatly, the length averaging $3\frac{1}{2}$ inches (8.9 centimeters), although it has been found to vary from $\frac{3}{4}$–9 inches (1.9–22.9 centimeters). Its position often varies from the normal so that it has been reported in every possible situation in the abdomen, depending on the position of the cecum and the length of the organ and its attachments.

An appendix is found only in humans, the higher apes, and the wombat, and possibly in some rodents. In herbivorous animals the cecum attains a very large size and it is thought by some that the appendix represents the degenerated remains of the herbivorous cecum. Others believe that it is a lymph organ functioning as do other lymph glands in the body. By many it is considered to be in the process of gradual obliteration in the human species. It is subject to inflammatory processes because of its limited blood supply, and since it is a blind tube, it is subject to obstruction from fecal impaction. See also **Appendicitis.**

APPETITE (Loss). See **Anorexia.**

APPLE LEAF SKELETONIZER (*Insecta, Lepidoptera*). The adult of this species (*Psorosina ammondi*, Riley), overwinters as a small dark-brown moth. Eggs are laid in early spring. The larva is a caterpillar that feeds on the underside of leaves and which later constructs shelters on the upper surface of leaves by drawing leaves together with silken webbing. The leaves become a mass of webbing and grass and are damaged by suffocation. The insect is distributed throughout the United States, but is most abundant in the central latitudes. The caterpillar is brownish-green and about $\frac{1}{2}$-inch (12 centimeters) long, with four shining black tubercles on back just to the rear of the head. This is not regarded as a serious pest because it occurs infrequently. The insect is mainly controlled by predators and parasites. When infestation is small, handpicking is effective. When infestations are large during the growing season, chemical controls, such as lime, may be used.

APPLE REDBUG (*Insecta, Hemiptera*). A sucking insect that attacks apple, haw, and pear and which occurs east of the Mississippi River in the United States and in southeastern Canada. The adult is orange-red with dark markings, up to $\frac{1}{4}$-inch (6 millimeters) long. The nymph is bright red, somewhat smaller than the adult. This insect, *Lygidea mendax* (Reuter), punctures the fruit, causing spots and deformation. These insects pass the winter in the egg stage. The eggs are laid in the bark of branches of trees and in the bark pores. They hatch in early spring. When light to moderate infestations are found during the season, cultural precautions for the next season may be the best practice. This involves using a dormant fruit tree oil spray that will kill the overwintering eggs. Treatment is similar to that for the aphid. See **Aphid or Plant Louse.** Where the infestation may be heavy during the growing season, control chemicals may be effective. Nicotine sulfate spray is also effective if applied at the cluster-bud stage.

APPLE TREE. See **(Rose Family).**

APPROXIMATE CALCULATION. It is often necessary, especially in physical problems, to make approximate, rather than exact, calculations. Typical cases include: the solution of polynomial or transcendental equations; evaluation of integrals which cannot be given in terms of simple functions; solution of differential equations with assigned boundary conditions.

Sometimes graphical or mechanical methods are convenient and these, while limited in their precision, frequently yield results of complete satisfaction. Numerical methods are more generally applicable, for with computers an answer can usually be obtained with any reasonable number of significant figures desired. In these cases, the procedure can be classified as one of two general classes: polynomial approximation or iteration.

The literature of this subject is very large and only a brief survey of it can be given here. Moreover, the methods involved can be modified in many different ways and there is no generally accepted name given to the various procedures.

Approximating methods also are described under **Curve Fitting; Differentiation (Numerical); Integration (Numerical);** and **Interpolation.**

APTERYGOTA. The primitive wingless insects with varying number of paired appendages on the abdomen and metamorphosis slight or absent. A subclass made up of the orders *Protura*, *Thysanura*, and *Collembola* in which the existing species are wingless and there is no evidence to show that wings have occurred in any ancestral form.

AQUACULTURE. This group of disciplines and practices, fundamentally centuries old in basic concept, essentially parallels agriculture (land crops and plants) and animal husbandry (livestock) in the quest to increase productivity of natural food substances, while simultaneously achieving efficiency and convenience of harvesting by way of concentrating the sources into smaller spaces. Possibly, the most useful synonym for aquaculture would be *fish farming* (where the term fish is interpreted broadly to include mollusks and crustaceans). Aquaculture can be broken down into several categories, usually in terms of environments and species involved, but possibly the most meaningful breakdown is (1) those operations directed toward the culture and harvesting of freshwater fishes; and (2) those operations directed toward marine animals, from which the terms *sea farming* or *mariculture* have developed. There are always the inevitable complications that detract from a crisp categorization— for example, the application of aquaculture to estuarine and brackish waters involving species that may be fully or partially tolerant to both fresh and saline waters. While most aquaculture operations are fully land-based or operate in the tidal zone between land and open water, there are also cultural operations that take place in naturally open and large bodies of water, such as lakes or various locations in the seas and oceans.

Bardach et al (1972) defined aquaculture as "the farming and husbandry of freshwater and marine organisms." Glude (1977) called aquaculture "the culture or husbandry of aquatic animals or plants by provate industry for commercial purposes or by public agencies for augmenting natural stocks." Brown (1977) uses the term "*sea ranching*—[a method that] involves releasing hatchery-reared animals of various sizes into marine waters for rearing and the subsequent recapture of the adult fish upon their return to the point of release." The general term, fish culture, is used by Edmunds and Lillard (1979).

Records indicate that aquaculture was practiced several thousand years ago in the Mediterranean area. It has been established that the Romans were successful in cultivating fishes in the brackish waters off the coast of Italy. Various Egyptian art forms depict the pond culture of fish. Records of the Chou and Shang Dynasties in China include reference to raising fish in ponds, projects of interest to the ruling class. Pioneered by Wen Fang in Honan Province (circa 1135–1122 B.C.), a treatise on these practices appeared in the Chinese literature as early as 475 B.C. To this day, aquaculture continues important to the food economy of China (Dill, 1967).

But, as pointed out by Sandifer and Smith (1978), despite its antiquity, aquaculture is a relative newcomer to the business of food production compared with agriculture and conventional fisheries. Consequently, aquaculture accounts for less than 1% of food production (Wheaton, 1977). Nevertheless, total food production from aquaculture exceeded 6 million metric tons in 1975. As aptly pointed out by Moiseeve as early as 1973, it is not the current levels of production, but rather the large potential productivity (hundreds of millions of metric tons) that is of such interest to a protein-short world population. For some of the developed countries, increased protein from aquaculture is not yet in serious contention, but some authorities are quick to point out that aquaculture may be required within the relatively near future if the desire for seafood luxuries, such as shrimp and other shellfishes, continues to develop at the pace enjoyed during the last few decades. In some countries, during the period 1960 to 1976, the demand for shrimp increased by about 80%, as compared with increased shrimp landings of less than 65%.

Compared with its relatively small beginning and slow development for many years, during the past decade fish culture in the United States has expanded steadily. Reasons given are high fish prices, the realization that conventional sources of fish are not limitless, the increased costs of catching fish in conventional fashion, and problems of pollution etc. that have interferred with natural fish habitats. As the vital need of the developed nations for increased energy sources becomes even more critical, trade-offs favoring energy will increase—as, for example, where historical and natural fisheries may be adversely affected by thermal and chemical pollution, among other factors. Aquaculture may provide an effective means in some instances for essentially abandoning traditional fisheries where there is a serious conflict with energy production.

The Worldwide National Oceanic and Atmospheric Administration estimates that 10% of the total fish production (1977) was derived from aquacultural operations, whereas this figure for the United States was about 3%, with the activities concentrated mainly on salmon, oyster, catfish, trout, shrimp, and clam species. Predictions by authorities as to the potential growth are mixed, ranging as high as 20 million tons (United States) by the mid to late 1980s, with the accompanying creation of many new jobs.

Much of the research on aquaculture in the United States has been directed toward determination of optimum growth conditions, to the formulation of rations which provide maximum feed conversion rates, to a reduction in the time required to produce a marketable product, and toward closing the life cycle of the shellfish when held in captivity. Edmunds and Lillard (1979) report that recent developments in closing the life cycle for shrimp and polyculture are technological advances which dramatically increase the economic feasibility of aquaculture for the affected species. Among other observations, these researchers found that cultured shrimp were equal or superior to wild shrimp in sensory quality. Their general findings ran counter in many instances to observations that cultured fishes of the various species are frequently inferior to the wild catches. References relating to the potential of aquaculture in the United States, as listed, include: (Anon. (1) through (5); Rutherford, 1975; Shoemaker, 1975; Gallese, 1976; Twitty, 1976; and Booda, 1976).

For many years, a number of countries have developed their sea resources in coastal regions. These estuarine regions have supported the culture of fish and invertebrates in tremendous numbers as well as great fisheries. In weight yield per surface area, these estuaries are much richer in food resources than the open sea.

Indian Ocean Regions. There are extensive culture fisheries in the estuarine areas of the Indian Ocean coastal zone. These brackish water swamps are banked with dikes to take in water at high tide, and the suspended silt settles on the bottom, slowly raising its level. Within several years, the land is raised sufficiently to be used as rice paddies. When the land level is high enough, intake of tidal water is discontinued and rainwater leaches the salt from the soil. After 2 to 3 seasons, the paddy may be planted. The plot will have a canal system inside the dike; these canals are used for culturing brackish water fishes. But, during this silting period, the area is used for growing fish. The tidal water brings in the fry of commercially important fish, including prawns. Screens prevent the escape of fish or the ingress of extraneous fish, and sluice gates control the tidal flow.

The once incidental use of swampy land for raising fish promoted a system of intensive brackish water fish culture in India and Pakistan. Fish ponds of different design and shape began to be constructed for the sole purpose of commercial fish culture.

On the islands of Java and Madura (Malaysia), ponds or "tambaks" produce upwards of 40 million pounds (18 million kilograms) of fish annually. The industry is based upon the *milk fish* (*Chanos chanos*). See Fig. 1. Post larvae and juveniles are collected from inshore areas from September to December and from April to May and transported to the fish pond areas in flat, watertight bamboo baskets. The annual requirements for the tambaks of these islands is estimated at upwards of 200 million fry.

Fig. 1. Milk fish.

In India, grey mullet, pearl spot, prawns, or milk fish are seined from nearby areas and transported to fish ponds called "porong" ponds. Each porong-type farm has fry ponds of varying areas, ranging from about 1000 to 10,000 square feet (90+ to 900+ square meters), and rearing ponds of 10,000 to 48,500 square feet (900+ to 4500+ square meters). The irregularly shaped sections are connected by secondary sluice gates and the whole complex is controlled by a main gate located in a deep portion, having a channel in the middle. The ponds are drained and dried to eradicate predatory and weed fish and to hasten decomposition of organic matter. Then from 1.2 to 2 inches (3 to 5 centimeters) of tidal water is taken in and allowed to stand. Within 3 to 7 days, a brownish-greenish-yellowish layer of microorganisms (mainly bacteria, unicellular and filamentous blue-green algae, and diatoms) develops on the bottom. Growth and production of fish in the ponds depends upon the growth of algae. The product is marketable within 6 to 10 months. The ponds are drained at low tide to capture the fish. The tendency of the fish to swim against the current is used for partial fishing. When tidal waters are let into the ponds, the fish swim against the current and then can be led to a catching pond.

In Java and Madura, annual production approaches 25 million pounds (11.3 million kilograms) of milk fish; about 10 million pounds (4.5 million kilograms) of penaeid prawns; and 6 million pounds (2.7 million kilograms) of other fish.

The sometimes-called giant Malaysian prawn (*Macrobrachium rosenbergii* de Man) is widely distributed in the Indo-Pacific region, ranging from Australia to New Guinea to the Indus River delta (Johnson, 1960; George, 1969; Malecha, 1977). There also have been reports of populations from the Palau Islands (McVey, 1975). In their review of the aquaculture of Malaysian prawns, Sandifer and Smith (1978), observe that adults are usually found in freshwater reaches of rivers, and during the breeding season, mature females migrate downstream into estuarine areas where the eggs are hatched. Presumably, the larvae are then swept further downstream into more saline waters where they develop. But, after a few weeks after the larval phase, the postlarvae begin to migrate toward freshwater where they grow to adults. Sandifer and Smith have diagramed the life cycle of the Malaysian prawn. See Fig. 2.

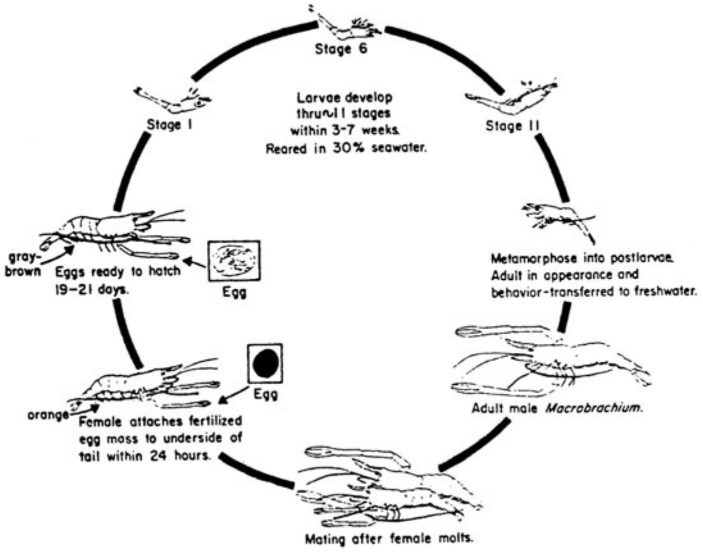


Fig. 2. Life cycle of Malaysian prawn. (*After Sandifer and Smith.*)

Controlled-environment experiments on cultivating Malaysian prawns have been conducted in recent years by the Marine Resources Research Institute, Charleston, South Carolina and these are reported in the Sandifer and Smith reference. The outdoor climate in this area permits prawn cultivation during only the warmer half of the year. Experiments, however, have shown that one crop of marketable prawns can be reared annually in ponds in South Carolina. It has been observed that a broad area of the state's coastal plain may be suitable for prawn farming. The researchers observe that to become commercially attractive, very intensive culture systems must operate continuously at high efficiencies and produce much greater yields per unit of area than a conventional pond system. External heat is required (possibly solar) for year-round operation. Culture containers (above ground) in form of tanks may be a limiting factor. Earthen ponds are being studied. Other configurations, including Shigueno-type tanks and aquacells are under consideration. Summarizing current problems, the researchers observe prawn nutrition, behavior, physiology, and genetics; culture systems design and optimization; pilot plant demonstrations; and economic feasibility analyses. A successful controlled-environment culture of *Macrobrachium* will likely require (1) the development of more nutritionally complete, cost-effective rations; (2) the determination and maintenance of optimal conditions for prawn survival and growth under crowded conditions; (3) genetic manipulation to produce prawns better suited for intensive culture than the essentially wild animals available today; (4) reduced system costs; (5) improved management techniques; and (6) greater production efficiency.

Mosquitofish (Nothobranchius taenipygnus), a tiny fish a little less than 1 inch long (2.5 centimeters) at the age of 3 to 4 weeks has been suggested by some authorities as a protein source in Africa and Asia. The fish will live in a few inches of water. Areas from 6 to 12 inches (15 to 30 centimeters) allow the maximum number to be bred in the minimum volume of water. Young *Nothobranchius* feed on algae and protozoa, but older fish prefer insect larvae. Apart from some chemical fertilizers to increase the growth of algae in the feeding pools, little has been done to provide supplies of feed. It is interesting to note that in 1977, researchers of the Sutter-Yuba Mosquito Abatement District in California reported on a formulation of a specific floating feed designed for rearing of mosquitofish. It is the plan to breed these fish in sufficient numbers to release them in California rice paddies for controlling mosquitos. This is an aquaculture project with a possible double advantage.

China. An excellent summary of aquaculture in China is provided by Ryther (1979). It is estimated that freshwater fish production in China probably exceeds that of the rest of the world combined, perhaps by as much as severalfold. Various estimates have placed the total area of freshwater culture in China at about 25 million acres (10 million hectares). Yields from the smaller, more intensively managed fish ponds have been estimated to range from a low of 893 pounds per acre (1000 kilograms per hectare) to 8930 pounds per acre (10,000 kilograms per hectare), with a probable average of about 2680 pounds per acre (3200 kilograms per hectare). It has been reported that the Ching Po Fish Farm (near Shanghai) produced 5000 pounds per acre (5600 kilograms per hectare) in 1978, with an objective of reaching 6700 pounds per acre (7500 kilograms per hectare).

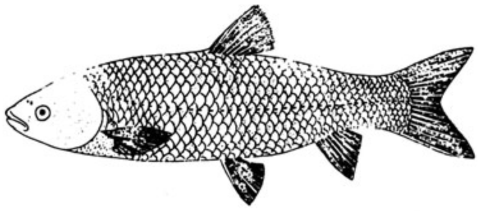
It has been estimated that the combined yields from aquaculture and marine fishing in China are equivalent to a per capita catch consumption of about 45 pounds (20.4 kilograms) against a backdrop of perhaps a billion people or more.

Ryther reports that the success of aquaculture relates to a number of factors: (1) multiple use of aquatic resources where several species are cultured in the same facility; (2) fish pond culture in China always has been considered an integral part of agriculture. Emphasis is given in polyculture to those species that are low in the food chain. Most of the fish species favored in Chinese pond culture are cyprinids (minnows), of which the 4 species (carp) illustrated in Fig. 3 are currently the most important. As stated earlier, it must be stressed that aquaculture in China dates back thousands of years.

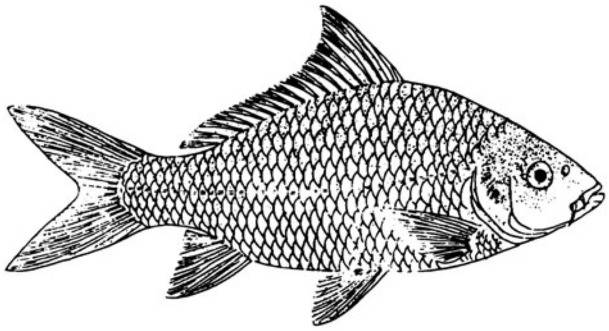
By contrast, marine aquaculture or mariculture represents a much more recent endeavor, extending back only a comparatively few decades on a large scale. Culture includes species of sea plants as well as sea animals. Probably the most important cultivated marine organism is the brown seaweed (*Laminaria japonica*), a cold-water species of kelp. See also the entries on **Algae;** and **Seaweeds.** Other cultured species include sea cucumbers, penaeid shrimp, pearl oysters, crab, mullet, and milk fish, but all are not in a fully developed state ready for commercial production. Mariculture has lagged somewhat, possibly largely because of the heavy accent on freshwater aquaculture, which is intimately tied to agriculture and a water conservation program. Aquaculture technology in China is headquartered in the City of Tsingtao, in which are located the Institute of Oceanography (Chinese Academy of Sciences), Shantung College of Oceanography, and the Yellow Sea Fisheries Institute (National Bureau of Fisheries). A significant part of the current research is directed toward algae and seaweed production.

Europe. In Italy, the Volturno River discharges into Lake Patria. The communication of the estuary with the sea consists of a very narrow shallow watercourse. Nutrients enter the lake from volcanic subsoil. Lake Patria teems with mullet and eels. The soft, muddy subsoil makes mussel farming on the bottom impossible, but by hanging cultures above the bottom, the mussels grow and fatten rapidly.

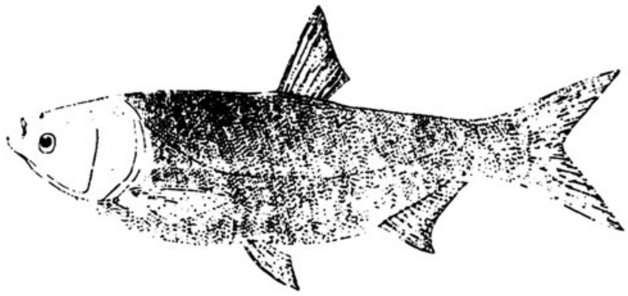
The Oosterschelde, in the southwestern section of the Netherlands, is an estuary with a different pattern. The Oosterschelde is an embayment penetrating far into the land that receives little influx from the surrounding fertile arable land. Yet river and seawater meet under the influence of the tides. The Rhine, the Meuse, and Scheldt Rivers discharge fresh water rich in nutrients and organic materials into the North Sea. Tides and winds bring about a thorough mixing and microorganisms mineralize the remains of freshwater organisms and other organic material. It is this mixed coastal water that pours into the Oosterschelde with the tides. The high and constant salinity, together with the discharge of nutrients conducive to rich plankton development, makes good oyster water. Under natural conditions, the oyster population was limited by an unprepared bottom and by failure of the oyster larvae to settle on beds where growth and fattening would be optimal. An exploitation of the natural resource by applying techniques of cultivation has resulted in a 30-fold increase in production over what nature formerly yielded.



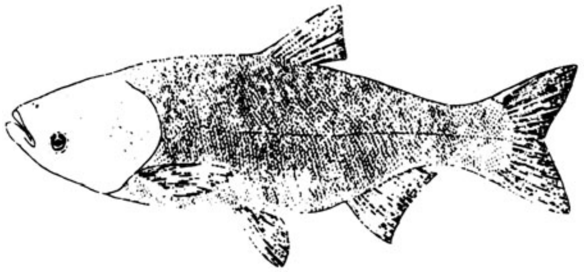
Grass carp *(Ctenopharyngodon idellus)*



Common Carp *(Cypnnus carpio L.)*



Silver carp *(Hypophthalmichthys molitrix)*



Bighead carp *(Aristichthys nobilis)*

Fig. 3. Four species of carp. (*S. Ling.*)

The Galacian bays of Spain are examples of estuaries in which shelter is a more important factor than discharge of nutrient-rich freshwater, since the seawater is sufficiently rich through upwelling. In the early 1950s, experiments were carried out to take advantage of the favorable conditions in these estuaries by growing mussels in hanging cultures following experiences gained in the Mediterranean Sea. Instead of racks, which are unsuitable where the tidal range is great, rafts were used. At first these were old boats equipped with outriggers. The success was promising and led to some 1500 installations consisting mainly of specially constructed rafts. Each raft carries 800 ropes; each rope is about 20 feet (6 meters) long and carries 100 pounds (45.4 kilograms) of marketable mussel. It takes eight months to rear a mussel seed to marketable size. See also **Mollusks.**

The Japanese use a similar technique to cultivate oysters. By growing oysters on long ropes hanging from simple rafts, the plankton of the whole water column is available to the oyster, and there can be a 50-fold increase in yield.

Another type of estuary, with extensive tidal flats and channels of varying depth through which tidal currents run, is located behind the Frisian Islands of Texel, Vlieland, and Terschelling, the Dutch Wadden Zee (western section). These flats are productive with mussels and cockles, brown shrimp, and polychaete worms of various species. Young flatfish and sole use this area as nursing grounds.

Several other examples of idealized situations for maritime culture could be described. Reference to McLeod (listed) is suggested.

Israel. Aquaculture commenced in Israel in the late 1930s and by the 1970s accounted for about 10% of the country's total annual value of agricultural production. Initial experiments utilized land and water unfit for ordinary agriculture. Productive ponds today are common throughout Israel with frequent integration of agricultural crops and fish ponds, all practicing required water conservation and reclamation. The Jordan River provides Israel with about one-third of its fresh water, the other two-thirds coming from ground water. Over 95% of the available fresh water is utilized, with emphasis being placed on use of brackish water in various agriculture/aquaculture practices. Several reasons for the development of an extensive fish culture industry in Israel include: (1) Skill in breeding and cultivation of carp was brought to Israel by immigrants from central Europe; (2) the need to provide dietary protein complicated by shortages of fresh water and good agricultural land; and (3) the eastern Mediterranean fisheries are not very productive, and projected declines in catches placed additional emphasis on expansion of intensive fish farming enterprises. In Israel, by the early 1970s, 8 pounds (3.6 kilograms) out of the 22 pounds (10 kilograms) of fish per capita consumed annually were pond-raised. The Fish Breeders Association is the governing body, controlling all aspects of freshwater fish production in Israel. All fish farmers are members of the association and pay a tax according to the area of their ponds and their production.

Fish raisers in Israel have found that only 4100 square yards (3400 square meters) of water area are required to produce one ton of fish, and portions of this can be used for irrigation. The basis of much of the technology is in controlled aeration of the pond area, thus permitting a two- to five-fold increase in fish stocking and the use of more sophisticated feeding methods. Yields of 2.8 to 12.1 tons per acre (7 to 30 tons per hectare) have been achieved by application of advanced technology as compared with yields of 1.2 to 1.6 tons per acre (3 to 4 tons per hectare) when using conventional procedures. The significant increase has been accomplished without a great change in the feed conversion ratio and total cost of production per ton. Other developments similarly have contributed to productivity of Israeli fish ponds. This has involved the use of supplemental feeding (25% protein pelleted feed), which has allowed an increase in pond stocking rate and improvement of tilapia yield in polyculture without depressing carp growth rates. Secondly, establishment of an optimum stocking density for both tilapia and mullet in the carp ponds has permitted more efficient exploitation of the pond's natural productivity.

High yields have been achieved through the use of multispecies (polyculture) culture. In polyculture, a range of fish species with a variety of feeding habits are stocked together, thus improving both the use and efficiency of conversion of the food.

The principal fish bred in the pond is carp, which feeds on grain and food wastes. Tilapis (Saint Peter's fish) and gray mullet are bred together with the carp. As noted, since each requires a different type of food (and occupies separate aquatic tropic levels), maximum use of land and water is achieved.

The species *Tilapia aurea*, introduced into carp ponds, originally presented a problem of wild spawning and resultant overcrowding and small size. Earlier investigations with tilapia were directed to finding ways to control unwanted spawning in the fish ponds. Prolific reproduction and reproduction at small sizes by tilapia results in gross overcrowding and consequent stunting of their growth. Lines of research aimed at controlling spawning included use of monosex culture (one sex of a fish present in a pond) and cultivation of tilapia in full-strength seawater where they can survive, but do not spawn. Another approach is the use of a predator fish (sea bass) that does not reproduce in fresh water and will control wild spawning.

Other cultured fishes in earlier stages of development include trout as well as (*Macrobrachium rosenbergii*), the Malaysian prawn. There is also early development of desert aquaculture in the lower Negev, including the integration of brackish pond aquaculture with intensive agricultural production in regions of high temperature and critical water availability. Cultivation of fish, such as tilapia, with variable-salinity tolerances, as well as adapted marine species, offers potential for polyculture in desert areas in other parts of the world, as well as in Israel.

Specific Aquaculture Objectives. Whether freshwater or marine species, the objectives of various countries as regards aquaculture cover a rather wide range. As of the early 1980s, these objectives in terms of species are given in the accompanying table. Only the main targets are given; the table is not meant to be all-inclusive.

Shortcutting the Conventional Food Chain. The small size of most of the food in the sea is one of the most serious constraints imposed by nature. In the ocean, unlike the land, most of the food is cycled through all but the final portions of the food chain in microscopic steps. The reasons for this are obscure, but they probably related to the absence over most of the ocean of a substrate in which large plants can attach at depths shallow enough for them to receive the light required for photosynthesis.

Each of the microscopic steps up the marine food chain is perhaps only 10 to 15% efficient, so that of the great amount of potential food initially fixed by photosynthesis in the phytoplankton, only a very small portion emerges in the form of fish to be caught.

The codfish, for example, eats predatory crustaceans and gastropods and small fish, which in turn, eat small herbivorous invertebrates. Thus, each million weight unit of cod production requires 10 million weight units of smaller fish and crustaceans, which in turn, requires 100 million weight units of small herbivorous invertebrates. These, in turn, consume 10 times their own weight in plant matter and a vast amount of bottom land.

McLeod (1969) estimates that about 70% of the total annual food requirements of the adult winter flounder are met by detritus and phytoplankton-eating invertebrates. Through a series of calculations, McLeod estimates that to produce 100 weight units of winter flounder, a total of 4088 weight units of direct and indirect vegetable matter is required.

For this reason, large-scale, controlled fish farms can be achieved only in restricted lagoons and estuaries, where the migratory propensities of genetically improved fish can be curbed, and where fertilization of the water with essential nutrients is accomplished by artificially induced upwelling or by direct introduction, and where both nutrients and fish can be retained.

In shallow waters, rice paddies, or estuaries, the autotrophic (plants) and heterotrophic (animal) layers are in close contact. In the sea, the autotrophs are small and the heterotrophs are large, whereas on land, the autotrophs are large (trees) and the heterotrophs are small. Organic detritus is the chief link between these levels of primary and secondary productivity, rather than a grazing food chain as on land. Although these productive areas may be less stable than the land, the advantages in such a detritus food chain is that microbial manipulations can be used to produce protein, which is available as food. Thus, an estuary has food all the time for its populations.

Certain steps in the food chain of the sea can be circumvented to increase the ultimate yield by a factor of 6 to 10 for each step circumvented. A filter-feeding creature, like a mussel, clam, or oyster, can

WORLDWIDE AQUACULTURE ACTIVITY

Country	Principal Species Cultured or Under Study for Potential Culture
Australia	Rainbow trout; crayfish (marrons and yabbies); native fish (Murray cod, golden perch, silver perch, catfish)
Austria	Rainbow trout; carp; tench; grass carp; lake whitefish
Belgium and Luxembourg	Rainbow trout; brown trout; carp; tench; roach; pike; pike perch; European eel
Canada	Salmon; trout; char
China (People's Republic of)	Common carp; crucian carp; grass carp; bighead; seaweeds
Denmark	Rainbow trout; brown trout; European eel
France	Rainbow trout; brown trout; carp; European eel; coho salmon; crayfish
Germany	Rainbow trout; brown trout; carp; European eel; tench; northern pike
Hungary	Carp; barbel; sturgeon; tench; bream; Danubian wels; pike-perch; pike; asp; brown trout; perch
Indonesia	Milk fish; common carp; grass carp; silver carp; kissing fourami; nilem; sepat siam; catfish; tilapia; shrimp
Ireland	Rainbow trout; brown trout; European eel
Israel	Carp; tilapia; mullet; silver carp
Italy	Rainbow trout; European eel; black bullhead; carp; mullet; sea bass; gilthead
Japan	Rainbow trout; freshwater eel; common carp; crucian carp; auy or sweetfish; loach; mullet; yellowtail; red sea bream (red porgy); horse mackerel; shrimp
Korea (South)	Carp; Japanese eel; rainbow trout; loach; grass carp; catfish; blue gill; bass
Netherlands	Roach; pike-perch; carp; pike; rainbow trout
Norway	Atlantic salmon; rainbow trout
Papua New Guinea	Rainbow trout; brown trout
Philippines	Milk fish
Portugal	Rainbow trout; brown trout; carp; steelhead trout; Atlantic salmon
Spain	Rainbow trout; Atlantic salmon; European eel
Sweden	Rainbow trout; Atlantic salmon
Switzerland	Rainbow trout; brown trout; carp; northern pike; lake whitefish; steelhead trout; char; pike-perch; perch
Taiwan	Milk fish; tilapia; Japanese eel; gray mullet; shrimp; carp
United Kingdom	Rainbow trout; carp; Atlantic salmon; turbot; Dover sole
United States	Rainbow trout; catfish; American eel; prawns; crayfish; salmon
Russia	Sturgeon; trout; carp; Pacific salmon

aggregate phytoplankton in one step into protein. Instead of moving uneconomically large masses of water to extract the plankton, efficient creatures can be used to concentrate the plankton wherever the natural movement of the water is sufficient to keep replenishing the plankton food supply. The concentration of plankton by natural advection is important in many places in the sea—at boundaries between current systems and around islands and shoals—and this can be a highly efficient concentrating mechanism. It can increase the local supply of plankton by a factor of 50 or more over the basic productivity of the contributing water.

Introduction of New Species. Closely allied to the technology of aquaculture is the introduction of new species. Reasons for introducing "new" species into a region include the market for a given species that may have to be imported from distant places to the establishment of a new industry and thus bolster the regional economy. In an excellent summary of the introduction of exotic species in aquaculture, Mann (1979) cites a number of examples, including (1) introduction of the Japanese oyster to the Pacific coast of North America and to France; and (2) introduction of the *Tilapia mossambica* (See Fig. 4), a native freshwater fish of Africa, to a number of countries, including Indonesia, Malaysia, Hong Kong, the Philippines, Vietnam, Cambodia, Laos, Taiwan, and China. The latter introduction was so successful that the fish is now one of the most important pond-cultured fish in southeast Asia. Infrequently, introductions of species can produce catastrophic instead of beneficial results and thus much trial and experimentation must go into planning such ventures. To date, as mentioned by Mann, the most disastrous introductions have occurred in terrestrial systems—rabbits in Australia, Dutch elm disease in the United States, European rats in the Pacific islands, and the giant snail, which has spread from West Africa throughout much of Asia. Guidelines for future introductions of aquatic species are being worked out by a number of organizations, including the International Council for the Exploration of the Sea (ICES).

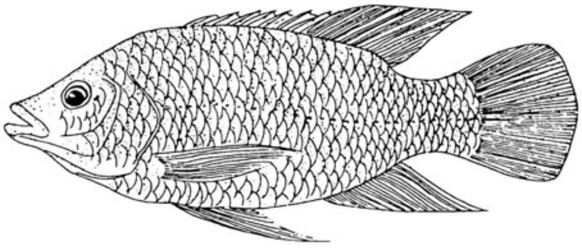


Fig. 4. Native freshwater fish of Africa. (*Tilapia mossambica.*)

Major Aquaculture Ventures in the United States. *Salmon.* Historically, the domestic cultivation of salmon in the Pacific Northwest has been a major aquaculture venture in the United States. Much technical attention has been directed toward maintaining salmon productivity. A new turn was taken in late 1970, when a pilot project was commenced to evaluate the rearing of Pacific salmon under captive conditions. With the cooperation of the National Marine Fisheries Service (NMFS), a joint government-industry[1] pilot program was initiated. Farms are located in the Puget Sound area of Washington. They involve two freshwater sites and two saltwater sites which sit across from each other at Manchester and Bainbridge Island, Washington.

Friedman (1978) points out that the Pacific coho salmon is an anadromous fish. It spends its early life cycle within freshwater and is not transported into saltwater until it possesses the physiological ability to osmoregulate within the saltwater environment. At this time, the

[1] Source: Union Carbide Corporation.

coho salmon from a morphological standpoint has also lost its vertical or parr marks and has obtained a silvery coat which represents the deposition of guanine and hypoxanthine.

The coho salmon spends the following 6 to 12 months in saltwater, developing to approximately 340 grams. At this time, the salmon may be harvested as a pansized fish, or permitted to mature. If allowed to mature, these salmon then become brood stock, are spawned directly from saltwater, and are utilized as a source of future progeny. The new season commences with the return of the adult salmon to freshwater to spawn. The coho orient against the most rapidly flowing portion of a stream and will readily enter a boxlike trap wherein they are captured and retained until sufficient quantities are obtained for spawning. See also **Salmon.**

In the process, once sufficient coho salmon have been gathered within the trap, they are checked for maturity, killed, and segregated by sex, body size, and conformation. Eggs are then removed from the two ovarian sacs of the female. Milt previously artificially stripped from male coho salmon and collected within a bucket immediately contacts the micropore of the egg and enters. After fertilization, water is added to the bucket, and the excess milt is washed away and the eggs are permitted to "harden" prior to movement into the hatchery.

The eggs then are incubated. Healthy eggs under incubation are normally pink to reddish-orange and somewhat transparent. Inspections are frequently made and dead eggs are removed to prevent fungal growth. Three weeks after hatching the small coho fry will have absorbed their yolk sacs and are ready for placement within freshwater pools. They remain there until they have the physiological capacity to undergo smoltification, that is, the ability to osmoregulate in saltwater. The young salmon are vaccinated and transferred to saltwater rearing pens. Throughout the history of the pilot project, bag-type net pens supported by flotation have been a key element in the salmon-culturing program. The present net pen system can produce over 660,000 pounds (300,000 kilograms) of protein annually. To feed the salmon a total of seven feed-storage bins are servicing sites and can store approximately 300,000 pounds (136,000 kilograms). The saltwater diets are formulated by linear programming techniques.

Catfish. One of the more recent ventures in aquaculture in the United States has been that of catfish farming.

Catfish farmers use either ponds, raceways, or cages. Three species of catfish are produced—the *channel catfish* (*Ictalurus punctatus*), the most common species raised; the *blue catfish* (*I. furcatus*), sometimes confused with the *channel catfish*; and the *white catfish* (*I. catus*). There are also bullheads (*Ictalurus* sp.), which are more difficult to manage and they often overpopulate ponds and can be quite troublesome, as well as bringing lower prices.

AQUAMARINE. A form of gem beryl. See **Beryl.**

AQUA REGIA. Also known as nitrohydrochloric acid, aqua regia is made up of three parts hydrochloric acid and one part nitric acid, each of the usual concentrated laboratory form. Aqua regia will dissolve all metals except silver. The latter is converted to silver chloride. The reaction of metals with nitrohydrochloric acid typically involves oxidation of the metal to a metallic ion and the reduction of the nitric acid to nitric oxide. Aqua regia also dissolves the common oxides and hydroxides of metals with the exception of silver, the ignited oxides of tin, aluminum, chromium, and iron, and the higher oxides of lead, cobalt, nickel, and manganese, the latter dissolving effectively in hydrochloric acid alone.

AQUARIUS (the water bearer). A relatively large constellation, important solely because of the fact that it is the eleventh sign of the zodiac. It contains no bright or particularly striking features. There is a theory that this constellation received its name because it appears over the Euphrates valley during that region's rainy season.

AQUEDUCT. An artificial conduit built to carry water is called an aqueduct. Generally speaking, aqueducts are built to convey a fresh water supply to congested districts from suitable sources more or less distant, and are therefore peculiar to cities. The first settlers in a place may depend upon local springs, streams, and wells, but with the growth of population there comes a time when these will prove inadequate, and suitable distant water supplies may have to be trapped through the medium of the aqueduct. An aqueduct may be either a pressure or grade-line type. Pressure conduits are commonly employed for small capacities or adverse topography, open or grade-line conduits for large capacity or favorable topography. Circumstances may require both types on the same project as the most economical combination. Pressure tunnels can convey water at pressures considerably above atmospheric, and are constructed with curved cross-sections. They are most frequently found in tunnel sections cut through hills and mountains, and in siphons. The principal distinguishing hydraulic characteristic of the pressure aqueduct is that it may depart from the normal open flow line both above and below the normal hydraulic gradient. However, siphon action, for sections above the hydraulic gradient, should be avoided wherever possible. Grade-line sections of aqueducts are usually built with open cut and fill construction following a hydraulic grade-line which will yield the requisite flow in the aqueduct at approximately atmospheric pressure, i.e., the fall per mile being just sufficient to overcome the friction loss in the same distance.

Some of the most important of the ancient aqueducts were those supplying the city of Rome, among which might be mentioned the Marcian, with a length of 58 miles (93 kilometers), the Julian, a length of 17 miles (27 kilometers), and the Claudian, with a length of 43 miles (69 kilometers). These were high level aqueducts of the grade-line type, principally cut and fill where possible, with grade-line tunnels for piercing hills, and resting on multiple arches when spanning valleys. These older aqueducts rarely had cross-sections greater than 30 sq. ft (2.8 sq. meters) in area. The Catskill aqueduct which conveys the water of the Ashokan Reservoir to the City of New York, approximately 100 miles (160 kilometers) away, has a capacity of 500,000,000 gallons (18,925,000 hectoliters) a day and is a splendid example of modern engineering on a large scale. It has in places cross-sections greater than 150 sq. ft. (14 sq. meters) in area, inverted siphons, one going more than 1000 feet (300 meters) below sea level, as well as a score of tunnels.

Irrigation projects throughout the world have required aqueducts and other kinds of waterways to transport immense quantities of water over long distances. The California Water Plan, originally put in place over a half-century ago and comprising a system of dams and waterways, is admired by agricultural and hydrological engineers worldwide.

AQUEDUCT OF SYLVIUS. The portion of the central canal of the nervous system of vertebrates which lies in the mid-brain, connecting the third and fourth ventricles.

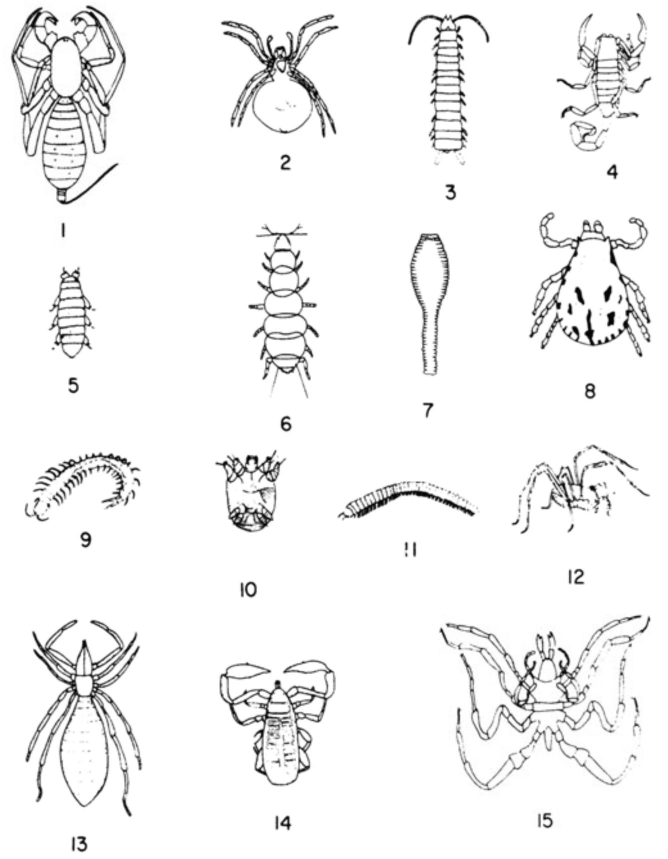
AQUIFER. See **Hydrology.**

AQUILA (the eagle). A constellation lying in the Milky Way, and hence containing rich star fields when viewed with a low-powered telescope. The distinguishing feature of this constellation is the group of three stars almost in a straight line, with the bright star Altair (α Aquila) between two fainter ones. Several novae have appeared in this constellation, the most famous one being Nova Aquilae III of 1918. See map accompanying entry on **Constellations.**

ARACHIDIC ACID. Also known as eicosanoic acid, formula $CH_3(CH_2)_{18}COOH$. A widely distributed, but minor component of the fats of certain edible vegetable oils. Shining, white crystalline leaflets; soluble in ether, slightly soluble in water. Sp gr, 0.8240 (100/4°C); mp 75.4°C; bp 205°C (1 millimeter pressure). Decomposes at 328°C. Commercial product derived from groundnut (peanut) oil. Used in organic synthesis, lubricating greases, waxes, and plastics. Source of arachidyl alcohol. See also **Vegetable Oils (Edible).**

ARACHNIDA. A class of the phylum *Arthropoda* including the spiders, mites, ticks, scorpions, pseudoscorpions, whip scorpions, sun spiders and harvestmen. Next to the insects this class is probably the best known among the invertebrates.

Arachnids differ from the other members of the phylum in one or more of the following characteristics: (1) The body is usually divided into two regions, a cephalothorax and abdomen. (2) Only simple eyes are present. (3) There are no antennae. (4) The thorax bears four pairs of legs in the adult. (5) The abdomen is often unsegmented and bears no appendages. (6) The first pair of appendages are chelate grasping organs. (7) Respiration is carried on by tracheae or lung-books. Several arachnids and related forms are shown in the accompanying illustration.



Arachnids and related forms: (1) Whip scorpion; (2) female black widow spider, *Latrodectus,* ventral view; (3) symphylid; (4) scorpion; (5) tardigrade or water bear; (6) pauropod; (7) linguatulid or tongue worm; (8) tick, *Dermacentor;* (9) centipede; (10) itch mite, *Sarcoptes;* (11) millipede; (12) opilionid; (13) solpugid; (14) pseudoscorpion; (15) pycnogonid or sea spider.

Arachnids are almost exclusively terrestrial and are predominantly predacious or parasitic, although some of the mites are plant feeders and the harvestmen include vegetable materials among their food.

The development of poison glands is fairly general in the group. Spiders have such glands, opening in the jaws or chelicerae, and scorpions have a special sting at the tip of the abdomen. With the exception of the black-widow spider of the United States and a small scorpion found near Durango, the poison is not known to be harmful to man. There is some probability that these two species may sometimes inflict wounds that are fatal if untreated.

The secretion of silk by spiders is another salient feature of the group. Silk glands are located in the abdomen, discharging through a group of spinnerets near the posterior end of the body. The silk is used to build webs of various forms for snaring prey, for the construction of cocoons to receive the eggs, as a lining for burrows, and in some cases as a vehicle to carry the animal on currents of air.

The economic importance of arachnids is rather limited. Spider silk has been woven but it is too delicate for extensive use and is valuable only as a source of cross-hairs for optical instruments. Aside from the poisons mentioned above, the principal harm from these animals is derived from the mites and ticks. The ticks do some damage to humans and domestic animals by sucking blood but their greatest damage is due to the transmission of diseases. Texas fever of cattle, Rocky Mountain spotted fever of humans, and other diseases are so conveyed. Mites living in the hair follicles, the sebaceous glands, and the tissues of the skin cause such diseases as scab in sheep and itch in humans. Plant-feeding species of economic importance include the bulb mite, the pear blister mite and various gall mites.

The classification of the arachnids is briefly as follows:

Order *Scorpionida*. The scorpions. Abdomen divided into a preabdomen and a slender postabdomen bearing a claw-like sting at the tip.

Order *Pedipalpi*. The whip-scorpions. Anterior pair of legs slender and antenna-like.

Order *Solpugida*. The sun spiders. Head and thorax separate.

Order *Chelonethida*. Pseudoscorpions. Very small scorpion-like animals; no postabdomen nor sting.

Order *Phalangida*. Harvestmen or daddy longlegs; commonly regarded as spiders but have a segmented abdomen broadly joined to the thorax.

Order *Araneina*. Spiders. Abdomen unsegmented and joined to the thorax by a slender waist.

Order *Acarina*. Mites and ticks. Small to moderate species with a sac-like body showing no well-marked divisions.

ARAGONITE. The mineral aragonite is calcium carbonate, $CaCO_3$, chemically identical with calcite but crystallizing in the orthorhombic system, with acicular crystals. By repeated twinning, pseudo-hexagonal forms result. Aragonite may be columnar or fibrous, occasionally in branching stalactitic forms called flosferri (flowers of iron) from their association with the ores at the Carinthian iron mines. Its hardness is 3.5–4; specific gravity, 2.93–2.95; luster, vitreous to resinous; colors, white, gray, green-yellow or purple; transparent to translucent. Aragonite forms at temperatures of 80–100°C and is relatively unstable at ordinary temperatures and pressures. It alters to calcite, although very slowly. There are many localities for aragonite in Europe, Bolivia, Pennsylvania, Iowa, Missouri, South Dakota, New Mexico, Arizona and Colorado. Its name is derived from Aragon in Spain. See also **Calcite.**

ARAUCARIAS. Trees of the family *Araucariaceae*, sometimes referred to as the Chile pine family or Pacific pines. These are evergreen trees of usually large size and heights ranging from 80 to 200 feet (24 to 60 meters) for adult trees. Their leaves are oval, leathery, and spirally arranged. The cones are round or oval. They are most unusual trees of the southern hemisphere. The Chile pine or monkey puzzle tree (*Araucaria araucana*) rises to a height of about 80 feet (24 meters). The leaves are long, pointed, triangularly shaped and of a yellow-green color. The tree is apparently immune to known diseases and can tolerate all but the wettest or driest of soils. A group of these trees tends to purvey the appearance of a quickly assembled Hollywood set for a jungle movie. The Parana pine or candelabra tree (*A. angustifolia*) is found in southern Brazil and rises to a height of about 110 feet (33 meters). It displays a limited number of branches, has a flat crown, and long hanging branchlets. The leaves possess curved spiny tips and are scaly. The bunya-bunya tree (*A. bidwillii*) occurs in Queensland and attains a height of 150 feet (45 meters). The cones are large, up to 12 inches (30 centimeters) in length. The Norfolk Island pine (*A. excelsa*) rises to a height between 150 and 200 feet (45 and 60 meters). The leaves are small, usually overlapping on older shoots. The cones range from 3 to 4 inches (7.5 to 10 centimeters) in length and are about as broad.

It is postulated by some that a bridge between the aforementioned and most unusual trees and the Giant Sequoias is filled by the Chinese fir (*Cunninghamia lanceolata*) on the one hand and by the cryptomerias or Japanese cedars. The Chinese fir occurs in central and southern China and is a large, spreading, and weeping evergreen that rises to a height of about 150 feet (45 meters). The crytomerias are of several species, including the aforementioned Japanese cedar (*Cryptomeria japonica*), which is a narrow, conical tree (older trees are dome-topped) and rises to a height of about 150 feet (45 meters). The Chinese fir is an important timber tree in China, as is the cryptomeria in Japan.

In the United States, the champion araucaria selected by the American Forestry Association is located in Honolulu, Hawaii. It is a Cunningham (*A. cunninghamii*) tree with the following dimensions: circumference at $4\frac{1}{2}$ ft (1.4 meters) above ground level of 189 in (4.8 meters); and a height of 120 ft (36.6 meters).

ARBOR. A shaft or stud, usually cylindrical or conical, on which a cutting tool, a tool holder, or a part to be machined is mounted or held.

ARBORVITAE. Trees and shrubs of the genus *Thuja*, family *Cupressaceae* (cypress family). The thujas include the large eastern white cedar (*Thuja occidentalis*) and the western red cedar (*T. plicata*). Although *T. occidentalis* can attain heights in excess of 100 feet (30 meters), the tree is frequently used in landscape planning and for hedgerows. The leaves are bright green, having overlapping scales. The cones are quite small. The natural range of the species is from southern Labrador to Nova Scotia, and west into Manitoba and Minnesota, and ranging south into Pennsylvania and following the mountains into North Carolina and eastern Tennessee. Specimens in the south are considerably smaller and may be described as small trees or shrubs. The species is quite common in northern New England. Asian thujas include the Japanese arborvitae (*T. standishii*), rare outside of Japan, and the Chinese thuja (*T. orientalis*), a small, often multistemmed tree. The *T. plicata*, also known as the zebra-striped western red cedar, is also used extensively as an ornamental tree.

In the United States, the champion arborvitae selected by the American Forestry Association is located in Maryland. It is an oriental (*T. orientalis*) or Chinese thuja with the following dimensions: Circumference at $4\frac{1}{2}$ ft (1.4 meters) above ground level of 100 in (2.5 meters); a height of 59 ft (18 meters); and a spread of 14 ft (4.3 meters).

ARC BACK. This is the occurrence of an arc from anode to cathode in a gaseous rectifier tube. Normally such a tube has electrons flowing from the cathode to the anode but under certain conditions excessive heating of the anode, excessive voltage across the tube, or other effects may cause the anode to emit electrons and allow an arc discharge to take place in a direction opposite to the normal direction. Under many circuit conditions this may destroy the tube or it may merely open the protective devices.

ARC CUTTING. Cutting of metal by means of an arc formed between the metal and an electrode. See also arc welding under **Welding.**

ARC (Electrical). A low-voltage, high-current electrical discharge, as contrasted with a spark. The electric arc, so called because of the shape of the "flame," was discovered by Davy about 1808. It is a type of discharge between electrodes in a gas or vapor which is characterized by a relatively low voltage drop and a high current density. The two types which are of considerable practical importance are the arc in open air and the arc in gases at low pressure. The familiar carbon-arc and the electric-arc furnace are examples of the former. In this type the arc is started by impressing a voltage across the electrodes in contact and then separating them. At the instant of breaking contact the high field and current density initiate the arc. Thereafter, if the current is kept constant, the potential necessary is a linear function of the interelectrode distance. In its steady state, the arc has an intensely hot cathode which emits a plentiful supply of electrons. The energy for heating the cathode is obtained from the high current density and from the bombarding positive ions. The arc is thus apparently a thermionic phenomenon. If the carbons are impregnated with a volatile metallic salt, the result is a "flaming arc," useful in producing the arc spectrum of the metal. If a dc carbon arc is placed in parallel with a suitable condenser and an inductance, the circuit so formed may be made to oscillate, as discovered by Duddell, and to serve as a source of undamped electric waves.

The arc is one of the most serious problems in switching electrical circuits, since the separation of the switch or circuit breaker contacts establishes an arc which must be extinguished in order to break the circuit. Many schemes have been developed to accomplish this. See **Circuit Breaker.**

The mercury-arc tube is the most important example of an arc in a gas at low pressure. Here the gas is the mercury vapor, the cathode is the mercury pool, and the anode is usually carbon. The arc is initiated by breaking contact between the mercury pool and a starting electrode. In gases at low pressure an arc may be established without breaking contact between the electrodes if the impressed voltage is high enough. If the voltage impressed across the two electrodes separated by a low-pressure gas is gradually increased the current increases slowly at first due to the residual ions and electrons present in the gas. The electrons ionize the gas molecules upon colliding with them, thus giving rise to additional ions and electrons to carry the current and also to produce more ionization by collision. This process continues until suddenly the breakdown voltage is reached, when the current increases very rapidly, even if the voltage is lowered. This is the initiation of the glow discharge. If the circuit resistance does not limit the current the discharge progresses almost instantaneously into an arc discharge having the distinguishing features of the regular arc. The discharge in thyratrons and other gas-filled hot cathode tubes is often called an artificial arc discharge since it is characterized by low voltage and high current density. It is not a true arc however since the cathode heat energy is supplied by an external source and not by the discharge itself.

ARC FURNACE. A furnace heated by an electric arc. The arc may be formed either between an electrode and the metal charge to be heated, or between several arcs placed over the charge.

ARC (Geometry). See **Circle.**

ARCH. An arch is a curved beam or truss made of wood, masonry, concrete or steel, whose supports are able to exert lateral as well as vertical forces to resist the action of any applied loads. These lateral forces are in the nature of thrusts which act inwardly toward the center of the arch span. The curvature of the beam must be in an upward direction in order to develop lateral reaction forces which will act in the required direction. The highest point of the arch is called the crown. The juncture of the arch and its foundation is called the springing.

In an arch, loads induce both bending and direct compressive stress. Although all loads are vertical, arch reactions possess horizontal components. Also, deflections have both horizontal and vertical components.

The masonry arch was used by the Egyptians, Babylonians, Assyrians, and Greeks but received full expression in the bridges and buildings of the Romans. Such arches depended upon the compressive thrust of adjacent blocks, or *vissoirs*, for their stability. That is, the line of pressure, or pressure line, had the same shape as the arch. Under the heavy moving loads of today the pressure line can depart appreciably from the arch axis and thus large tensile forces due to bending can be developed. Masonry is unsuited for such conditions and it is necessary to use reinforced concrete or steel arches. Steel arches are of the solid-rib or trussed type.

These structures may be further classified as fixed or hinged arches. A fixed arch is a structure which is rigidly connected to its supports in such a manner that they exert vertical and lateral reactions and prevent rotation. A two-hinged arch is one which is free to rotate about its supports, consequently it is able to exert only vertical and lateral reactions. Three-hinged arches have an additional hinge, usually midway between the hinges at the supports. The tied arch is a structure in which the lateral forces are applied by means of a horizontal tension member connecting the ends of the arch.

From the standpoint of structure, an arch is similar to a rigid frame. See Fig. 1. Arches are not commonly precast because stacking for transportation is difficult due to the curvature. However, small spans cast at the site have been used successfully. Prestressing seldom is advantageous because arches are subjected to large compressive forces.

Arches are frequently used for very long spans and support roofs of

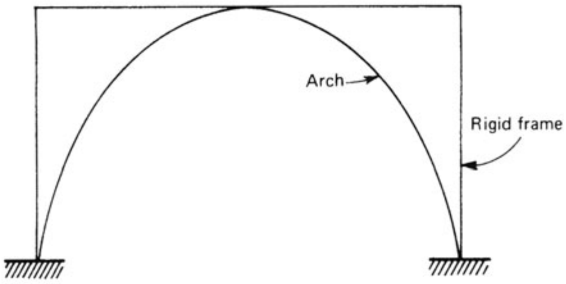


Fig. 1. Comparison of arch with rigid frame.

such structures as auditoriums, exhibition halls, hangars, stadiums, sporting arenas, and transportation terminals. Typical arch bridges are shown under **Bridge (Structural).**

One of the most interesting arches is the 630-foot (192-meter) high Gateway Arch, made of stainless steel, situated in downtown St. Louis, Missouri. This catenary arch was designed by Eero Saarinen and is the tallest monument in the United States. See Fig. 2.

Fig. 2. Gateway Arch in Saint Louis, Missouri. Structure is 630 feet (192 meters) high.

ARCHEGONIATES. Those plants in which the female sex organ is an archegonium are sometimes called Archegoniates.

ARCHEGONIUM. The multicellular female sex organ characteristic of all the *Embryophyta* except the Angiosperms. It consists of a swollen basal portion called the venter, and an elongated neck. The venter may be a single layer of cells, but is often many cells thick. Within the basal portion is contained the single large egg, and a second, somewhat smaller, ventral canal cell, while the elongated neck contains a single row of cells which eventually dissociate, leaving an open canal through which the sperm may pass to reach the egg.

ARCHEOCYTE. Cells of sponges which ingest and digest food, carry the products to other parts of the body, and form reproductive cells. They are amoeboid (see **Amoeba**) and are found in the mesenchyme.

ARCHEOLOGICAL DATING. See **Mass Extinctions; Radioactivity.**

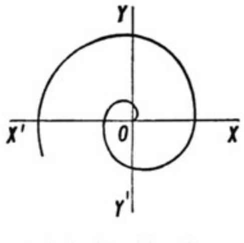
ARCHEOPHYTIC. A paleobotanic division of geologic time. See also **Paleobotany.**

ARCHEOZOIC (Archean). The oldest of the five eras of the earth's history. The rocks of this system are metamorphosed equivalents of all types of sedimentary and igneous rocks, but principally the latter. No undisputed fossils have been found in the Archean. The lower Archean (Keewatin) of North America is composed of a preponderance of metamorphosed, basaltic lava flows and tuffs with some metamorphosed sediments, such as quartzite and slate. The general character of the basal Archean proves that the oldest known rocks do not represent the original crust of the earth. The upper Archean (Laurentian) contains a preponderance of granite, gneisses, and schists in the form of batholiths intruding the Keewatin. The principal areas of Archeozoic rocks are in Canada, Finland, Scandinavia, Australia, Africa and northeastern South America. Many of the formations contain rich ore deposits, especially of gold and silver. Large amounts of graphite suggest the former existence of life. Length of time since the beginning of the Archeozoic is possibly more than 2500 million years. Owing to the lack of fossils, structural complexity, high degree of vulcanism and metamorphism, geologists have found great difficulty in deciphering the history of this earliest recognizable portion of the "crust" of the earth. The structural history of the Archean is therefore not so well known as that of the succeeding periods and intercontinental correlation is particularly difficult. On the other hand, the search for ore deposits has been an important stimulus to the study of the Archean formations, especially in Canada.

A highly generalized section, about 25 miles (40 kilometers) long, showing the relations of the Archeozoic group of rocks in the Lake Superior-Lake Huron region of Canada. The Keewatin system was moderately folded and intruded by the Laurentian granite, after which there was deep erosion. Then the Timiskaming rocks were laid down, and later, were strongly folded and intruded by the Algoman granite, after which there was another period of profound erosion, marked by the upper surface.

ARCHIANNELIDA. A class of annelid worms including small marine species of simple structure.

ARCHIMEDES SPIRAL. A special case of a spiral, represented in polar coordinates by $r = a\theta$. It may be described as the locus of a point moving with uniform velocity along the radius vector, while the

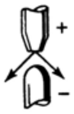


Spiral of Archimedes.

radius vector also moves about the pole with constant angular velocity. The evolute of this spiral approaches asymptotically to a circle with radius *a*.

Another spiral, similar to this one, with polar equation $r = a\theta^2 - b$, is known by the name of Galileo. See also **Hyperbolic Spiral.**

ARC LAMP. The electric-arc lamp has, as its source of illumination, an electric arc struck between two electrodes. In contrast to the incandescent lamp, in which the illumination results from a heated filament, and vapor lamps, in which the illumination is derived from a vapor made luminous by electric current, the light from an arc lamp comes from the highly incandescent crater of one of the electrodes, and from the heated, luminous, ionized gases surrounding the arc. The light from arc lamps is very much more intense than that from the incandescent type lamp. From the standpoint of current consumption, the illumination is produced efficiently. Some arc lamps may not be operated on ac, but all types are adaptable to dc. A constant-current, series type circuit is used to operate street-light arc lamps.



Arc-lamp electrodes.

Open-arc type lamps can be divided into: (1) flame arcs; (2) low-intensity projector arcs; and (3) high-intensity projector arcs. The entire arc stream is made luminescent by the addition of flame materials in the flame arc light. This type of lamp is used in photography and related industrial photochemical processes because it produces an essentially continuous radiation and closely approximates natural sunshine. A special coring of rare-earth salts is used. In a special version of the flame-arc lamp, the arc is surrounded by a glass globe with limited access of air, thus resulting in a nitrogen-rich atmosphere. This enhances the radiation in the violet and near-ultraviolet regions, making the lamp useful in blueprinting and related copying processes.

The incandescent tip of the positive carbon electrode, at or near its sublimation temperature, is the main light source in the low-intensity projector arc. The brightness is uniformly generated over a considerable area. This type of lamp found early use in motion picture, searchlight, and other projector systems requiring a concentrated source of light and the ability to create a well-defined narrow beam.

In the high-intensity carbon arc, rare-earth materials are included in the core of the positive electrode. These materials volatilize into the arc stream as the electrode is consumed. The light from these lamps is well suited to color motion picture photography and projection.

ARC (Mathematics). A segment or piece of a curve. See also **Circle (Geometry).**

ARC SHOOTING. (1) A method of refraction seismic prospecting in which the variation of travel time with azimuth from a shot point is used to infer geologic structure. (2) A reflection spread placed on a circle or on an arc with the center at the shot point.

ARCTIC CIRCLE. The line of latitude 66° 32′ N (often taken as $66\frac{1}{2}$° N). Along this line the sun does not set on the day of the summer solstice, about June 21 and does not rise on the day of the winter solstice, about December 22. From this line the number of annual twenty-four hour periods of continuous day or of continuous night increases northwards to about six months at the North Pole.

ARCTIC REGION RESEARCH. See **Polar Research.**

ARCTIC WATERS. Water masses in or associated with the Arctic Ocean, including:

Arctic Deep Water. An oceanic water mass with a salinity of about 34.95% and temperature of about −0.85°C (30.5°F). Because of its high salinity, it is not believed that it is formed in the Arctic Ocean, but rather it is considered to be the North Atlantic Deep and Bottom Water, flowing north.

Arctic Surface Water. An oceanic water mass of low salinity, averaging from 32.0% to 33.0% in the north, but reaching very low values in summer as the rivers carrying large volumes of fresh water from melting ice decrease the salinity near the surface to values far below 30.0%. As winter approaches, the salinity increases, due to slow mixing with the Arctic Deep Water.

See also **Ocean.**

ARCTIC ZONE. Geographically, the area north of the Arctic Circle (66° 32′ N).

ARCTURUS (α *Bootes*). Arcturus was probably one of the first stars to be named, most likely receiving its name because of its proximity to the constellation of Ursa Major, thus being called the "watcher of the bear." It is one of the few stars named in the Bible (Job IX), although this reference is evidently to the constellation of Ursa Major rather than to the actual star itself. References to Arcturus are to be found in the writings of many of the ancient poets, including Virgil.

Arcturus is a very interesting star from the astronomical point of view, being what is known as a giant star. In appearance, it is a reddish-yellow, and the spectral type is such as to indicate that its temperature is slightly lower than that of the sun. It is one of the few stars whose diameter has actually been measured with the stellar interferometer. The diameter is about 30 times that of the sun.

Arcturus ranks fourth in apparent brightness among the stars and has a true brightness value of 110 as compared with unity for the sun. Estimated distance from the earth is 36 light years. Arcturus is classified as an orange star of spectral type K. See also **Constellations;** and **Star.**

AREA. If a parallelogram has sides denoted by the vectors **A** and **B**, its area is given by the vector product, $\mathbf{C} = \mathbf{A} \times \mathbf{B}$, which is perpendicular to the plane determined by **A** and **B.** The scalar magnitude of **C** equals that of the area and the direction of **C** is arbitrarily taken as the direction of the outward normal to the surface.

In calculus, the area of a surface may be found by integration. If $y = f(x)$ describes a curve, the area bounded by the curve, the X-axis, and the ordinates (a, b) is

$$\int_a^b f(x)\, dx$$

If a surface of revolution is generated by rotating an arc of the curve about the X-axis, its area is given by the integral

$$S = 2\pi \int_a^b y\, ds$$

where $ds^2 = dx^2 + dy^2$. A multiple integral may also be used, for an infinitesimal surface element in the XOY-plane is $dx\, dy$ and the area over a region S is

$$\iint_S dx\, dy$$

For a curved surface described by $z = f(x, y)$, the area is

$$\iint_S f(x, y)\, dx\, dy$$

Formulas for calculating the areas of major surfaces and shapes are given throughout this volume. See **Circle (Geometry).**

AREA SAMPLING. A method of sampling in which the domain to be examined is divided into small areas some of which are then selected to form the sample. Each area so selected may be fully inspected or may form the basis of sub-sampling. The method is particularly appropriate where no lists are available of the primary units which form the target population, e.g., individuals or dwelling units in developing countries. See also **Sampling (Statistics).**

ARENACEOUS. A textural term applied to sediments or sedimentary rocks which are composed of grains of sand. Psammitic has the same meaning.

ARÊTE. A narrow, jagged, serrate mountain crest, or a narrow, rocky, sharp-edged ridge or spur, commonly present above the snowline in rugged mountains (as the Swiss Alps) sculptured by glaciers, and resulting from the continued backward growth of the walls of adjoining cirques. (*Glossary of Geology*, American Geological Institute).

ARGAND DIAGRAM. A graphical method of representing a function of a complex variable, $z = x + iy$. There are two perpendicular axes, as in the usual rectangular Cartesian coordinate system. The real part of the function is plotted on the real axis, usually the horizontal one, and the imaginary part on the imaginary or vertical axis. Points on the diagram for various values of the number pair (x, y) can then be joined to give a curve for the function of the complex plane. See also **Complex Variable.**

ARGENTITE. The mineral argentite, sometimes called silver glance, is naturally occurring silver sulfide, corresponding to the formula Ag_2S. It crystallizes in the isometric system in cubes, octahedrons and dodecahedrons, or may be massive. Hardness, 2–2.5; sp gr, 7.2–7.34; luster, metallic; streak, gray; color, black, blackish-gray or gray; opaque and sectile to such an extent that it cuts like wax with a knife. Heated upon charcoal it yields a malleable mass of silver. The name is derived from the Latin word for silver, *argentum.*

Localities for fine crystals are Sonora, Mexico, and Freiberg, Saxony; in the United States, at Butte, Montana; Tonopah, Nevada; and Aspen, Colorado.

Argentite is probably the most important primary silver mineral. However, it maintains its cubic (isometric) characteristic only above 179°C (354°F). Upon cooling, the inward structure inverts to a non-isometric form, usually orthorhombic, yet retaining its original outward form. It is, therefore, a paramorph after argentite, known as acanthite.

ARGILLACEOUS. This term is used to designate sedimentary rocks composed of fine particles of the nature of clay or mud. Pelitic has the same meaning.

ARGILLITE. A dense, fine-grained, hard, sedimentary rock of various colors (usually white, gray or red). Composed of minute grains of both clay and quartz. Certain types of argillites are easily confused with certain types of fine-grained acid lava flows, such as felsites, unless studied microscopically.

ARGON. Chemical element symbol Ar, at. no. 18, at. wt. 39.948, periodic table group 18 (inert or noble gases), mp $-189.2°C$, bp $-185.7°C$, density 1.78 g/cm^3 (solid at $-233°C$). Solid argon has a face-centered cubic crystal structure. At standard conditions, argon is a colorless, odorless gas and does not form stable compounds with any other element. Because of its low valence forces, argon is unable to form diatomic molecules, except in discharge tubes. It does form compounds under highly favorable conditions, as excitation in discharge tubes, or pressure in the presence of a powerful dipole. As an example of the first, argon forms amorphous compounds of the type FeA in a discharge tube having iron electrodes. An example of the second is furnished by the hydrates which argon forms with H_2O at 150 atmospheres and 0°C. Argon forms compounds, possibly clathrates, with a number of organic substances, such as a compound with hydroquinone containing 9% argon, in which the amount of argon may vary from this proportion. The compounds are made by crystallization of the aqueous solution of the hydroquinone under argon gas pressure on the order of 40 atmospheres.

Argon occurs in the atmosphere to the extent of approximately 0.935%. In terms of abundance, argon does not appear on lists of elements in the earth's crust because it does not exist in stable compounds. However, argon is 2.5× more soluble in H_2O than nitrogen and thus is found in seawater to the extent of approximately 2800 tons per cubic mile (605 metric tons per cubic kilometer). Commercial argon is derived from air by liquefaction and fractional distillation. There are three natural isotopes, ^{36}Ar, ^{38}Ar, and ^{40}Ar, and four radioactive isotopes, ^{35}Ar, ^{37}AR, ^{39}Ar, and ^{41}Ar. The lengths of half-lives of the isotopes vary widely, the shortest ^{35}Ar with a half-life of about 2 seconds; the longest ^{39}Ar with a half-life of about 260 years. The first ionization potential of Ar is 15.755 eV; second 27.76 eV; third 40.75 eV. Other important physical characteristics of argon are given under **Chemical Elements.**

The presence of argon in air was suspected by Cavendish as early as 1785, but was not positively identified until 1894 by Lord Rayleigh and Sir William Ramsay. Argon exhibits a characteristic series of lines in the red end of the spectrum. Commercially, argon gas is used in incandescent lamps and fluorescent lamps as an inert gas to minimize vaporization of the filaments and, for this, is preferable to nitrogen. The gas also is used for shielding electrodes in arc welding. A gas of about 99.995% purity is required for lamps. Argon also has found effective use in certain lasers.

Regarding argon in meteorites, see **Krypton.**

Ar Isotopes in Paleochronology. Isotopes of Ar and their ratios have proven useful in geological period dating and thus are of assistance in establishing databases for paleoclimatology and estimating periods of mass extinction on Earth. For example, major changes in Earth's climate are known to have occurred during transition between the Eocene, q.v., and Oligocene, q.v., periods. This was a major transition from the essentially tropical environment of the Mesozoic, q.v., to the start of the galacial world. This is considered the most significant climatic alteration since the prior demise of the dinosaurs. However, it has been difficult to establish the timing of prehistoric events, for lack of evidence.

Since the mid-1960s, $^{40}K-^{40}Ar$ dates have been used to estimate time periods in terms of Ma (millions of years), and some adjustments were made to prior estimates of the Chadronian boundary (36–32 Ma). Other isotopic data (^{40}Ar, ^{40}Ar/^{39}Ar) and Rb-Sr have been used for time estimating.

As pointed out by Swisher and Prothero (1990), recent advances (1989–1990) in mass spectrometric techniques and the development of laser-fusion ^{40}Ar/^{39}Ar dating techniques have resulted in the ability to date individual volcanic crystals. Multiple analysis and the ability to date single crystals allow the identification of multiple-age components, due to detrital contamination, and thus permit improved precision and accuracy. To date, studies have been directed to North American chronology, notably minerals (biotite, anor, plag) found in Nebraska and Wyoming.

Additional Reading

Francis, A. W.: "Argon" in *Encyclopedia of Chemistry* (S. P. Parker, Editor). McGraw-Hill, New York, 1983.

Kent, J. A.: *Riegel's Handbook of Industrial Chemistry*, Ninth Edition, Van Nostrand Reinhold, New York, 1982.

Perry, R. H., and D. Green: *Perry's Chemical Engineers' Handbook*, 6th Edition, McGraw-Hill, New York, 1988.

Staff: *Handbook of Chemistry and Physics*, 73rd Edition, CRC Press, Boca Raton, Florida, 1992–1993.

Swisher, C. C., III, and D. R. Prothero: "Single-Crystal ^{40}Ar/^{39}Ar Dating of the Eocene-Oligocene Transition in North America," *Science*, 760 (August 17, 1990).

ARGONAUTA (*Mollusca, Cephalopoda*). The genus to which the paper nautilus belongs. This species is not a true nautilus but is more closely related to the octopus.

ARGUMENT. 1. An independent variable; e.g., in looking up quantity in a table, the number or any of the numbers which identifies the location of the desired value; or in a mathematical function the variable which when a certain value is substituted for it the value of the function is determined. 2. An operand in an operation on one or more variables.

ARGYRIA. Poisoning from the use of silver preparations over too long a period or its absorption during industrial processes, causing a ghastly bluish discoloration of the skin over the entire body.

ARIES (the ram). This constellation is far more famous for its classical significance than because of its appearance in the sky. In contains no bright stars and has no conspicuous features. Two thousand years ago, the vernal equinox was located in the constellation of Aries, and the symbol for the vernal equinox is the symbol for the constellation (i.e., the ram's head). Precession has caused the position of the vernal equinox to move backwards into the constellation of Pisces, so that now the "sign of the first of Aries" is to be found in that constellation. (See map accompanying entry on **Constellations.**)

ARIL. In many plants there is formed in the developing fruit an outgrowth from the funiculus, or seed stalk, one which completely or partially surrounds the seed. In the litchi nut it is the thick translucent pulp surrounding the seed; in the nutmeg it is a meshlike envelope which when removed and dried is known as mace.

ARITHMETIC. The rules for combination of two or more numbers. The operations involved are addition, subtraction, multiplication, division and the results obtained are, respectively, a sum, difference, product, quotient. Two or more of these operations can be performed successively and the final number obtained is independent of the order of the intermediate steps. Arithmetic, which is the first kind of mathematics normally studied by the beginner, is essentially the art of computation and application of this art.

See also **Progression** for meaning of arithmetic series and progression.

ARITHMETIC MEAN. The arithmetic mean of a number of observations x_i ($i = 1 \cdots n$) is the simple average $\Sigma\, x_i/n$. It is commonly referred to simply as the mean.

The arithmetic mean is the most generally useful measure of location. In samples from a normal distribution, the mean is an efficient, and indeed sufficient estimate of the location parameter, though in other distributions other estimators may be preferable. If $\bar{x}$ is the mean of a sample drawn from a finite population of N members, $N\bar{x}$ is an unbiased estimate of the population total.

If κ_p is the pth cumulant of the parent distribution, that of the distribution of the mean is κ_p/n^{p-1}; in particular, the variance of the mean is σ^2/n. As this result indicates, the distribution of means from any parent distribution satisfying rather general conditions tends to normality as the sample size increases. See **Central Limit Theorem.**

A useful generalization of the arithmetic mean is the weighted mean. If each observation x_i has attached to it a weight w_i the weighted mean is given by $\Sigma\, w_i x_i/\Sigma\, w_i$, observations with large weights being given greater influence. If we have a set of estimates x_i of a quantity ξ, and if the variance of x_i is σ^2/w_i with w_i known, the estimate of ξ with least variance is the weighted means of the x_i and the variance of this estimate is $\sigma^2/\Sigma\, w_i$.

ARKOSE. Arkose is a relatively coarse-grained feldspathic sandstone, derived from the rapid disintegration of granite or other feldspathic rock. It is characterized by its content of fresh, unaltered, euhedral feldspar. The term was proposed by Brongiart in 1823, and has been in constant use ever since. Arkose is an important type of sediment, especially in relation to the study of unconformities, paleoclimatology, and "fossil" soils.

ARM. An extended lobe or appendage of a body. Its most familiar use is in application to the pectoral appendages of vertebrates when freed from the usual functions of support and locomotion, as in humans and the other primates. In anatomy, the term is restricted to the region from the shoulder to the elbow, to distinguish it from the forearm. Some species are still quadrupedal on the ground, but can use their arms to some extent for handling objects.

The radiating lobes of the starfish are called arms or rays and arm is also applied to the branches of the lophophore in brachiopods and to other special structures.

Robots are also said to have arms (also hands, wrists, etc.). See **Robot and Robotics.**

ARMATURE. The armature is one of the two essential parts of the dynamo electric machine. In a generator, the armature is the winding in which electromotive force (emf) is produced by magnetic induction. In the motor armature, conductors carry the input current which, in the presence of a magnetic field, produces a torque and effects the conversion of electrical into mechanical energy. In dc machines it is the rotor, but the ac armature may be rotor or stator. Larger size synchronous machines always have stationary armatures. The reluctance of the magnetic circuit to the flux which the conductors of the armature must cut in order to generate electric energy, is decreased by providing a core of soft iron or steel, on the surface of which the conductors are embedded in slots suitably provided in the core. The armature windings of a dc generator are terminated at the segments of a commutator, by means of which the alternating emf's induced in the armature are rectified and transferred by brushes from the moving rotor to stationary terminals. The conductors must be separately insulated, as must be also the commutator segments, and must be well braced and anchored in their slots to resist the electromagnetic and mechanical forces which tend to displace them.

The term armature is also applied to the moving element of a magnetically-actuated relay.

ARMATURE REACTION. This term refers to the reaction of the magnetic field produced by the current flowing in the armature conductors upon the main magnetic field of a dynamo machine. The result is a distortion of the magnetic field, the extent depending upon the reluctance of the magnetic circuit, the arrangement of the armature windings, the type field structure, and the phase angle between the armature voltage and current. In dc machines the effect is to increase the flux at some pole tips and decrease it at others, while in ac machines the effect depends upon the field structure and the phase angle of the armature current and voltage. The flux may be distored as in the dc machine, it may be changed in magnitude but undistorted in wave form or it may be changed in magnitude and shifted in position with respect to the field windings. Armature reaction is an important factor in the speed and voltage regulation of the machines.

ARMYWORM (*Insecta, Lepidoptera*). An economically important caterpillar, *Cirphis unipuncta*, named from its habit of migrating from field to field in large numbers. When severe outbreaks occur these insects completely strip fields of grain of all kinds. When migrating they are trapped in barrier ditches dug around the fields to be protected. They are also killed by poison baits.

The true armyworm develops from a moth (*Pseudaletia unipuncta*) of the moth family *Noctuidae*. It is found in the United States and Canada, east of the Rocky Mountains. The eggs develop on the underside of leaves.

The armyworm is smooth, dark green with long white stripes extending from front to back, and often is fat and up to two inches (5 centimeters) in length.

Occurrence of the armyworm is subject to cycles, not fully understood or predictable. The insect is particularly damaging to maize (corn). In the young corn plants (8 inches (20 centimeters) or less in height), the armyworm devours all or nearly all of the leaves, causing the plant to expire in short order. In taller plants, some or parts of the leaves are left intact, but the insect concentrates on the center of the young stalk and damage is usually fatal to the plant—and once a field is invaded, without immediate control measures at work, all plants in the field are consumed.

The armyworm commences destructive action early in the spring. It is believed that some of these insects winter over as partially developed larvae, but that some may also winter as pupae or adults. This would explain the very early appearance of moths in the northern climes. Depending upon locale, the worms are fully grown by April or May. They pupate just below the groundlevel of the soil. A pupa is dark brown, having a length of about $\frac{3}{4}$-inch (19 millimeters), with a rather blunt head and a sharply tapering tail. The pupa stage lasts for about 2 weeks (longer if weather is unusually cold), from which the insect transforms into a rather drab gray-brown or light-brown moth having a wingspread of about 1.5 inches (about 4 centimeters). The moth can be recognized by a prominent white dot, centered in each front wing. The moth is active only at night. Light and the odor of decaying fruit attracts the moth.

The female moth lays her eggs (500 or more) in clusters or rows, preferring to place them on leaves of grasses. Pale green in color, the young worms tend to "loop" as they move about. These worms may be found by the thousands and tens of thousands in grass and fields of small grain and, frequently, are not detected until their damage has become extensive. The worms feed only at night. When fully grown, the worm is about 1.5 inches in length (4 centimeters). They maintain a green-brown color and have longitudinal stripes. After feasting for several days, the worms enter back into the soil, change to the pupal stage, and re-emerge as moths in from 2 to 4 weeks. In this fashion, there may be two to three generations per year. The larvae of the last generation of the year usually appear between mid-August and mid-September, depending upon locale and weather conditions.

Poison bran is an effective measure, particularly to protect fields which have not yet been attacked. The bran is placed in a line stretching across the probable line of march of the insect once an adjacent field has been decimated. The bait is prepared from bran that is mixed with insecticide. Blackstrap molasses or lubricating oil usually is added to provide sufficient stickiness to keep the insecticide and bran together. The worms also like molasses.

The armyworm is quite similar to the army cutworm, described in the entry on **Cutworm.**

A closely related species is the *fall armyworm (Laphygma frugiperda*, Smith). Although widely distributed, they are notably injurious to crops in the southern United States, particularly during years when there has been a cold and wet spring. Like the armyworm and the cutworm, these insects also prefer small grains, maize (corn), sorghum, and grasses, but they also attack alfalfa, bean, cabbage, groundnut (peanut), cucumber, potato, sweet potato, spinach, and turnip. These worms are particularly fond of lawn grasses and thus are a serious economic pest not only to the food producer, but to the homeowner as well.

Actually, the fall armyworm is considered a tropical insect because it cannot winter over in any area where the soil is frozen hard. Thus, a favorite winter ground is along the Gulf Coast and in southern Florida. Here, during winter, several stages of the insect may be present at the same time. After their number is increased many fold in the spring, they swarm northward, often flying many hundreds of miles before selecting a location for their egglaying. Each female moth lays about 1000 eggs, usually on green plants. The female covers clusters of eggs with hairs from her body. Shortly thereafter, the young larvae descend down through the heart of a plant and continue feeding near groundlevel until they assume a length of from 1 to 1.5 inches (2.5 to about 4 centimeters). It is at about this time that the insects are noticed as the result of the large amount of damage that becomes apparent. Unlike the armyworm and most cutworms, the worms do not take refuge in the soil during daytime, but rather they cling to parts of plants.

Quite similar in appearance to the true armyworm, the fall armyworm larvae when fully grown are a light-tan to green in color, although sometimes they are black. There are three very narrow white stripes down their back. They can be contrasted with the true armyworm by observing a white inverted "Y" design on the front of the head. Also, the tubercles are more prominent and they have more hair. The marching habit of this worm occurs during the autumn in the northern climes, but can take place in the southern states any time after the middle of summer and, if weather conditions are ideal for the insect, such marches may occur in early spring. An entire field or garden can be consumed within 36 to 48 hours. The remaining life cycle of the fall armyworm is similar to that of the true armyworm.

Another closely related species is the *beet armyworm (Laphygma exiqua)*. This is a large caterpillar ranging up to $1\frac{1}{4}$-inch (30–32 millimeters) in length when fully mature. It is olive-green with broad light-green striping. Sugarbeet is the favorite target crop of this insect, followed by table beet and a variety of vegetables, citrus, alfalfa, and some wild grasses. This insect, native to the Orient, was first noted in California in the late 1870s and now occurs widely in the Gulf States, and from those states westward into California and northward to Nebraska and Kansas. Its habits are quite similar to those of the fall armyworm. Control is similar.

AROIDS. A large group of monocotyledonous plants, mostly tropical, having a characteristic flower habit. The numerous small inconspicuous flowers are borne on a fleshy stalk or spadix, which is surrounded, more or less completely, by a large, expanded, often brightly colored bract called a spathe. The spadix and spathe together are often but incorrectly considered to be the flower of the plant. The aroids are perennial plants, generally having tubers or rhizomes from which rise large leaves. Many tropical members are climbing plants. Well-known species are the Skunk Cabbage, whose foul-scented flowers appear so early in the spring, the Jack-in-the-Pulpit, and the wild arum, *Calla palustris*, of cold swamps, as well as the Sweet Flag, *Acorus calamus*, of the marshes. The cultivated Calla Lilies are all aroids and not lilies at all; some of them are delightfully fragrant. On the other hand, in species of *Amorphophallus*, which are sometimes seen in collections of cultivated plants, the vile odor of the flower structure prevents them from becoming popular; the spathe and spadix of some of them are of gigantic size. In the tropics several species of *Colocasia* are cultivated for the edible rhizomes which appear under the name of dasheen or taro. *Monstera* and several species of *Philodendron* are popular decorative plants in homes and in public buildings.

AROMATIC COMPOUND. An organic compound that incorporates a closed-chain or (ring) nucleus in its structure. This is in contrast with the aliphatic compound which is comprised of an open-chain structure. The classical example of an aromatic compound is benzene. Aromatic compounds also are sometimes referred to as benzenoids. Some ring-type compounds are not classified as aromatic. These include the cycloparaffins and cycloolefins which are considered to be derivatives of methane. See also **Compound (Chemical); Organic Chemistry.**

ARRHENIUS-GUZMAN EQUATION. A relation between the viscosity η and temperature T, at constant pressure,

$$\eta = A \exp \frac{B}{RT}$$

where A, B are constants, and R is the gas constant; B may be identified with the *activation energy for liquid flow*.

ARRHENIUS VISCOSITY EQUATIONS. (1) Effect of temperature on viscosity, η, of a liquid

$$\frac{d}{dT} \ln(\eta v^{1/3}) = \frac{k_1}{T^2}$$

where v is the specific volume and k_1 is a constant.

(2) Viscosity of solutions, η,

$$\eta/\eta_s = A^x$$

where x is the concentration, η_s is the viscosity of the solvent and A is a constant.

(3) Viscosity of a sol, η,

$$\log \eta/\eta_\infty = kC$$

where $\eta_\infty$is the viscosity of the medium and C is the concentration of the sol-forming material.

ARRHYTHMIAS (Cardiac). An arrhythmia is a variation from what is considered normal for a rhythmic phenomenon. An analogy would be an erratic tape recorder whose speed vacillates and differs from that *one* speed which yields perfectly normal reproduction of music or voice. While the heart has a reasonably wide range in beating rate (pulse rate), depending upon the body's energy requirements (spanning from rest to heavy exertion), the rhythm of heart action is preserved even though rate may be changed. Rhythmic disturbances of the heart are called *cardiac arrhythmias*. Because, as explained in **Heart and Circulatory**

System (Human), the heart is comprised of chambers (sinuses) and valves which are governed by electrical impulses, the impulses in a normal heart are programmed in just the right sequential order and at just the right instant, the timing of which is on the order of milliseconds. A crude analogy is the firing order of a multi-cylinder internal combustion engine.

Some authorities classify cardiac arrhythmias as *passive* and *active*. In some cases, both passive and active arrhythmias may be simultaneously present. The detailed etiology of these conditions is beyond the scope of this encyclopedia. The principal tool available to the physician in the analysis and diagnosis of cardiac arrhythmias is the electrocardiogram. *Bradycardia* is the term usually used to describe an abnormal slowness of the pulse rate. *Tachycardia* is the term used to describe excessive rapidity of heart action. *Heart block* is a term that signifies the interruption of muscular connection between the atrium and ventricle so that they beat independently of each other. Since there are numerous electrically conductive pathways within the heart, there are several specifically named heart blocks. These include: *Arborization* heart block in which there is interference with the fine terminal subendocardial fibers of the *Purkinje system* (specialized sino-atrioventricular conduction system); *atrioventricular* heart block; *auriculoventricular* heart block; *bundle-branch* heart block, in which the two ventricles contract independently of each other; *complete heart block*, a situation in which the functional relation between the parts of the *bundle of His* (a muscular band connecting the auricles with the ventricles of the heart) is destroyed by a lesion, so that the auricles and ventricles act independently of each other; *interventricular* heart block; and *sino-auricular* heart block, in which the blocking is located between the auricles and the mouths of the great veins and coronary sinus. See Fig. 1 in entry on **Heart and Circulatory System (Human).**

Because of the several pathways of conduction, a pair of conducting leads to the electrocardiograph does not suffice. A standard arrangement of twelve leads to the instrument is commonly used. The physician will evaluate with maximum possible accuracy the risk posed by an arrhythmia, particularly a tachyarrythmia and the urgency of attempting to terminate it. Although normal hearts may tolerate tachyarrythmias for extensive periods, when there is marked underlying heart disease, such arrythmias may cause hypotension (low blood pressure), congestive heart failure, or coronary insufficiency. Tachyarrythmias are prone to degenerate into the more serious ventricular arrythmias. In addition to general cardiac health, the physician will take into consideration the heart rate, the duration of the episodes of arrythmia, and whether the arrythmia is regular or irregular.

Passive Arrythmias

Particularly in individuals who are quite physically active, *sinus bracycardia* may be present. This is defined as a resting heart (adult) which beats fewer than 60 times per minute. The rates of long-distance runners, for example, may range between 40 and 50 beats per minute. A number of conditions, however, may cause an abnormally slow discharge rate of the sinus node. This occurs during severe pain and can be induced by increased vagal tone caused by various drugs (parasympathomimetic), such as endrophonium and neostigmine, and by a number of tranquilizing drugs. Increased vagal tone also can be associated with acute myocardial infarction. A number of physiologic situations will slow the sinus node, including hypothyroidism and high fever. There is also what is known as *hypersensitive carotid sinus reflex*. (The carotid artery is located in the region of the neck, face, and skull.) The carotid sinus becomes increasingly sensitive with age and a type of syncope (sudden suspension of consciousness) can be caused by twisting the neck or wearing a tight collar. See also **Syncope.**

In some patients, particularly the elderly, presenting sinus bradycardia, there may be a condition known as the *sick sinus syndrome*. The bradycardia may be punctuated by episodies of tachycardia (*bradycardia-tachycardia syndrome*). A form of sick sinus syndrome has explained some sudden deaths of young athletes, who at autopsy have exhibited an idiopathic (cause not known) obliterative disease of the artery to the sinus node.

A slow heart beat (bradycardia) may be a factor in development of congestive heart failure with patients who have myocardial disease. See also **Congestive Heart Failure.**

Active Arrythmias

Some authorities place active arrythmias into two fundamental categories—supraventricular and ventricular. Representative of *supraventricular arrythmias* are atrial premature beats and junctional premature beats; sinus tachycardia; paroxysmal supraventricular tachycardia; paroxysmal atrial tachycardia; paroxysmal (AV) atrioventricular junctional tachycardia; paroxysmal atrial tachycardia with block; multifocal atrial tachycardia; atrial flutter; and atrial fibrillation. Representative of ventricular arrythmias are ventricular premature beats, ventricular tachycardia, accelerated idioventricular rhythm, and ventricular fibrillation.

Ventricular Fibrillation. This condition may be described as an irregular twitching of the muscles in the wall of the ventricle of the heart. About two-thirds of the deaths due to heart attack are attributed to uncontrolled ventricular fibrillation (VF) which kills many patients before they reach a hospital. The probability of VF increases with the size of the dead tissue (infarct) that results when the blood flow to a portion of the heart is drastically reduced by blockage of one or more of the arteries. This is a primary motivation for finding treatments to limit infarct size. The presence of dead tissue from prior milder heart attacks will contribute to VF in a major attack. Thus, in persons with a history of heart problems, surgery to eliminate abnormal tissues that may give rise to the arrhythmia is sometimes suggested as a preventive measure. When a patient lives to reach a hospital, continuous monitoring of the heart will reveal any abnormal rhythms that presage VF and thus can be treated as soon as they occur.

Various communities have instituted emergency procedures which provide *cardiopulmonary resuscitation* (CPR), including mouth-to-mouth resuscitation and external heart massage. Therapy must be commenced within 5 minutes to prevent irreversible brain damage when the supply of blood to the brain is blocked. A major role of CPR is to identify VF and the immediate use of drug therapy. As of the late 1980s, successful CPR programs require nearly an ideal environment to operate successfully, i.e., large numbers of trained personnel and sufficient units to reach a heart attack victim within 5 minutes or less; the prompt attention of bystanders (if attack occurs in public places), of family in the home, of fellow office and factory workers, etc., to take rudimentary measures and, above all, telephone for an emergency unit; adequate instrumentation, including transmission of instrumental data as well as voiced observations, in the emergency vehicle; excellent radio communication between vehicle personnel and hospital medical staff (a physician immediately available at all times) to advise most appropriate emergency procedures. The objective is to extend as much as possible the skills and means of the hospital to the patient before arrival at a hospital.

Preexcitation Syndromes. These conditions occur when all or a portion of the ventricle is activated by atrial impulses earlier than would be the case were the impulses reaching the ventricles via the normal conduction pathways. One such condition is known as the *Wolff-Parkinson-White* syndrome and is believed to result from congenital cardiac defects, frequently involving the tricuspid valve (*Ebstein's anomaly*). Recent research indicates that preexcitation syndrome can occur by a number of different pathways. Sometimes involved are the Kent bundles, the James fibers, and the septal fibers of Mahaim. Although the arrythmias (usually sporadic) associated with preexcitation syndrome can be tolerated by some patients (even at rates of 250–300 beats per minute), this tolerance is usually limited to young persons with otherwise normal hearts. In contrast, in some patients, the arrythmias can be disabling.

Pacemakers

In cases of complete or bifascicular heart block, an implanted pacemaker may be used. In the early years of pacemaker use, there were numerous problems with failures arising from electrode displacement or breakage, premature battery depletion, or faulty pulse generators. Today, the reliability of the units has markedly increased, but precaution must be exercised in selecting an instrument from the several types available. Two techniques have been used in pacemaker implantation—insertion of an electrode through (1) a cephalic vein or (2) the external jugular vein into the apex of the right ventricle. The power source is placed subcutaneously just inferior to the clavicle; or direct epicardial electrodes are attached to the ventricle through a small midline thoracotomy. The pulse generator is placed subcutaneously in the epigastric region. In the *demand type* of pacemaker, the unit will not compete with the patient's own rhythm if the AV conduction should return, or with ventricular premature beats should they be present. The pacemaker can be designed and implanted in accordance with the particular problems of the patient.

Over a half-million patients in the United States live with the aid of a permanent cardiac system to ensure a dependable cardiac rhythm, with the number increasing by several thousand each year. The latest available systems take advantage of the advancements in microelectronics. Such advancements have led to complex multiprogrammable single- and dual-chamber arrangements for pulse generation. But, as pointed out by Ludmer and Goldschlager (see reference), the large number of available pacing systems has complicated the understanding of newer generator functions and pacemaker electrocardiography, and thus the ability to diagnose normal and abnormal pacemaker function. The proper methods and techniques of outpatient follow-up also have been complicated.

The modern pacemaker generator is a hermetically sealed metal can weighing from 30 to 130 grams, powered by a lithium battery projected to last 2 to 15 years. The modern unit is noninvasively (transcutaneously) programmable in more than 42 million possible setting combinations. Noninvasive programming, by which certain pulse-generator functions can be transiently and reversibly changed, has revolutionized the pacemaker industry and avoids reoperation in up to 20% of patients.

The parameters that are reprogrammable in many modern pacemakers include: (1) rate, (2) energy output, (3) refractory period, (4) sensitivity, (5) hysteresis, (6) mode of function, (7) lower rate limit, (8) atrioventricular delay, and (9) upper rate limit. The refractory period of a pacemaker generator may be defined as the time after either a paced complex (pacing refractory period) or a sensed spontaneous complex (sensing refractory period) during which the generator is unreceptive to incoming signals. In some pacemakers, pacing and sensing refractory periods are equal, whereas in others, the refractory period after a paced complex is longer than that after a spontaneous complex.

The complex technology of pacemakers tends to rival the sophisticated instrumental techniques applied in controlling a complex chemical plant and thus suggests that those physicians who specify pacemakers today must understand and appreciate the relevant sophistication of these devices. For the patient who has interests along these lines, the Ludmer/Goldschlager reference should be "must" reading. Because of these complexities, this is an area of medical specialization where a second opinion may be highly justifiable. Two causes of problems with all pacing systems are the inability to sense electrical signals of borderline quality (undersensing) and the capability of sensing unwanted electrical signals originating from outside the heart (oversensing). Irrelevant signals may include those of a physiological nature, electromagnetic interference emanating from any other medical aids of an electrical nature, or those encountered at home or at work, such as radio and television transmitters and arc-welding equipment, and still another source of signals that may be generated within the pacing system itself, such as caused by inactive leads or electrode parts of the pacing device.

In another excellent reference (Phibbs/Marriott), the overall importance of the pacemaker in relieving patients of symptoms and in saving lives is fully acknowledged. The researchers also stress, however, that it would be wrong to pretend that the implementation of permanent pacing is not without some risk. Medical complications fall into three principal categories: (1) thrombosis and embolism, (2) infection, and (3) less common complications, including pacemaker-generated arrhythmias, myocardial perforation, and tamponade. The risk of serious thrombotic or embolic complications is estimated at approximately 2% of cases. A relatively benign form of thrombosis of the veins of the upper arm and shoulder may be expected in about 30% of patients with transvenous pacemakers. Thrombosis of the axillary and subclavian veins is much more serious, possibly requiring anticoagulant and thrombolytic therapy. Pulmonary embolism, often fatal, has been the subject of numerous case reports. Pulmonary embolism in a patient with a pacemaker should always arouse the suspicion of thrombosis on the pacing wire as the source. The dimensions of the problem have not been fully defined. Infection from pacemaker implantation can be expected in from 1 to 7% of cases, usually taking the form of septicemia or as endocarditis. Phibbs and Marriott report that patients with pacemakers encounter major difficulty in obtaining life and health insurance. Psychological problems associated with pacemaker patients are also reported.

Statistics have shown that many pacemakers in the past have been implanted needlessly in patients, as for example the use of a pacemaker to correct for a situation that was of a transient nature with little prospect of recurrence. Studies along these lines have been conducted at major hospitals and medical centers. Phibbs/Marriott report that in a few localities as many as 75% of pacemaker implantations were found unjustifiable by any reasonable standards after review by several disinterested experts. The total figure for larger areas is certainly much lower, but in some regions and states it approximates 30%. Causes for misdiagnosis and incorrect recommendation of a pacemaker include transient situations as previously mentioned, hypothyroidism, simple misinterpretation of an electrocardiogram and, among less obvious situations, the effects of various drugs on cardiac arrhythmias. Drug-induced depression of atrioventricular conduction is often confused with intrinsic dysfunction or disease. A specific example is given by Phibbs/Marriott—an elderly patient with chronic atrial fibrillation who has a slow ventricular rate while receiving a digitalis preparation. Instead of discontinuing digitalis or lowering the dose, a surprising number of physicians have implanted a pacemaker for this iatrogenic, completely reversible abnormality.

It is further observed that there are impelling medical and legal reasons for total removal of *unnecessary* pacing equipment. Approximately 50% of the older types of electrodes used and practically all of the newer, long-tined electrodes will become permanently entrapped. "In the large percentage of these patients in whom the pacemaker is entrapped, removal will require cardiopulmonary bypass, with major morbidity and possible death. Failing to attempt to remove an electrode as soon as it is found unnecessary could reasonably be construed as falling below an acceptable level of practice, since such failure may expose the patient to a life-threatening intervention."

Antiarrhythmic Agents

Some of the drugs used in the control of ventricular and atrial tachyarrhythmias include:

Quinidine—important in the control of ventricular and atrial tachyarrhythmias. The drug decreases automaticity, excitability, and conduction velocity. It prolongs effective refractory period and action potential duration. Although quinidine is essentially metabolized in the liver, several instances of drug interactions involving quinidine have been observed.

Procainamide—similar in action to quinidine. The drug is as effective as quinidine for treatment of ventricular tachyarrhythmias, but is not very effective in treatment of atrial tachyarrhythmias. Most common early noncardiac manifestations of procainamide toxicity include gastrointestinal disturbance, anorexia, nausea, and vomitting. The drug also can cause drug fever or allergic rash.

Propranolol—decreases conduction through the atrioventricular node. Often an effective agent for slowing the ventricular response in patients with atrial fibrillation or atrial flutter. The major toxic effects are related to its beta-blocking activity, which may precipitate excessive sinus bradycardia, congestive heart failure, or in patients with bronchospastic disease, bronchial asthma.

Lidocaine—effective in treatment of ventricular arrhythmias. Not effective in treatment of supraventricular tachyarrhythmias. Cardiac toxicity is observed as negligible at therapeutic levels. Major toxic effects are neurologic, including confusion, seizures, and (rarely) respiratory arrest and coma.

Mexiletine—a lidocaine derivative that provides oral therapy for ventricular tachyarrhythmias due to many causes. Side-effects may include epigastric burning and constipation. Neurologic toxicity may include tremulousness, diplopia, dizziness, and (infrequently) slurred speech.

Tocainide—a lidocaine derivative that provides oral therapy for ventricular tachyarrhythmias due to many causes. Side effects include nausea and vomiting.

Digitalis—stabilizes atrial electrical activity and assists in preventing atrial tachyarrhythmias. Particularly useful, in combination with diuretics, in treatment of ventricular arrhythmias arising from congestive heart failure and enlarged heart.

Verapamil—the first of the calcium antagonist drugs to be approved for clinical use in the United States, although previously used in other countries for several years in the treatment of arrhythmias and angina. The calcium antagonists inhibit the flux of calcium across the slow channels of vascular smooth muscle cells and cardiac cells. Adverse reactions have been reported when verapamil and propanolol are given intravenously (simultaneously).

Amiodarone—a benzofuran derivative with reported effective use for treatment of supraventricular and ventricular tachyarrhythmias. The drug has been used for a number of years in countries other than the United States. The drug, with a long half-life, may persist in the body for over two months after therapy has ceased. Amiodarone interacts with other drugs, potentiating the anticoagulant properties of warfarin and increasing the serum concentration of several other drugs. Side-effects from long-term therapy are considered mild, including skin reactions, photosensitivity, mild gastrointestinal effects. As with all drugs, there are uncommon manifestations.

For related topics in this encyclopedia, see list of entries given at the end of entry on **Heart and Circulatory System (Human).** Particularly, see **Electrocardiography.**

ARROW WORM. Small marine animals sometimes classified with the annelid worms but more often included in the separate phylum *Chaetognatha*.

ARROYO. This term is applied to dry stream channels with nearly vertical walls and flat bottoms which are characteristic of semi-arid regions. They may suddenly become filled with torrential waters after heavy rains. The word arroyo is of Spanish origin.

ARSENIC. Chemical element symbol As, at. no. 33, at. wt. 74.9216, periodic table group 15, mp 817°C (24 atmospheres), sublimes at 613°C, density 5.72 g/cm^3. One naturally occurring stable isotope ^{75}As. Various studies indicate that arsenic exists in several allotropic forms. The metallic form has a steel-gray color in the crystalline form and is brittle. Although the red form of arsenic sulfide As_2S_2 was observed by Aristotle as early as 400 B.C., the first attempt to isolate the metal was not made until 1250 by Albert Magnus. Later documentation on the preparation of the element was given by J. Schroder and N. Lémery in the 1600s. First ionization potential 9.8 eV; second 18.63 eV; third 28.34 eV; fourth 50.1 eV; fifth 62.5 eV. Oxidation potentials $AsH_3 \rightarrow As + 3H^+ + 3e^-$, 0.54 V; $As + 2H_2O \rightarrow HAsO_2 + 3H^+ + 3e^-$, −0.2475 V; $HAsO_2 + 2H_2O \rightarrow H_3AsO_4 + 2H^+ + 2e^-$, −0.559 V; $AsO_2^- + 4OH^- \rightarrow AsO_4^{3-} + 2H_2O + 2e^-$, 0.71 V; $As + 4OH^- \rightarrow AsO_2^- + 2H_2O + 3e^-$, 0.68 V. Other important physical properties of arsenic are given under **Chemical Elements**.

Gray or metallic arsenic, density 5.73 g/cm^3, which sublimes on heating, and has the vapor composition As_4, becoming As_2 at higher temperatures, is the ordinary variety. On rapid cooling, the vapor condenses to yellow arsenic, density 1.97 g/cm^3, which reverts to the gray variety on warming. An intermediate form in the transition is black amorphous β arsenic, density 4.6–5.2, also obtained by the thermal decomposition of arsine. Brown arsenic, density 3.7–4.2, obtained by reduction of acid solutions of trivalent arsenic, is probably a finely divided form of black arsenic.

Arsenic sublimes on heating; is unchanged in dry air but a film of oxide is formed in moist air; heated in air at 180°C forms arsenic trioxide of the odor of garlic, poisonous; insoluble in HCl but soluble in concentrated HNO_3 or concentrated H_2SO_4 to form arsenic acid; soluble in hot NaOH solution; heated with chlorine forms arsenic trichloride; heated with metals forms metallic arsenides. When arsenic is heated in a tube and the vapor cooled (1) slowly (that is, in the hot part of the tube) black arsenic is formed, and this form is converted into the gray at 360°C, (2) rapidly (that is, in the cold part of the tube) yellow arsenic is formed, and this form is quickly converted into the gray by the action of light. Yellow arsenic is soluble in CS_2.

Arsenic occurs in nature as the arsenide of iron, cobalt, nickel, and as the mineral sulfides, *realgar* (arsenic monosulfide, AsS), red colored; *orpiment* (arsenic trisulfide, As_2S_3), yellow colored—these two minerals when powdered once were used as paint pigments—*arsenopyrite, mispickel* (iron arsenosulfide, FeAsS); *enargite*, Cu_3AsS_4; and *tennantite*, $Cu_8As_2S_7$.

The primary arsenic-containing material is arsenious oxide obtained by separation from roaster or smelter flue gases. Metallic arsenic is obtained as sublimate by heating the oxide with carbon.

Arsine: AsH_3, is formed by hydrolysis of arsenides, or reduction (by zinc and HCl or aluminum and NaOH) of arsenic compounds, as in the Gutzeit test. It reacts with metals at higher temperatures or in solution to give the arsenides. Diarsine, As_2H_4, is produced by reduction of arsenic trichloride, $AsCl_3$, by lithium aluminum hydride, $LiAlH_4$ in ether at −190°C. It melts below −50°C, but begins to decompose into AsH_3 and brown polymeric $(AsH)_x$ about −100°C. It is more stable in the gas phase than in the solid or liquid phases.

Arsenides: These are prepared by fusion from the elements. Their properties vary across the periodic table, those of the alkalies and alkaline earths being readily hydrolyzed by H_2O or acids and are stoichiometric, while the arsenides of the other metals show an increasingly intermetallic character and resist hydrolysis.

Gallium Arsenide: Devices using gallium arsenide are extremely important in contemporary electronic equipment. For example, some of the fastest digital integrated circuits built to date are made of gallium arsenide. They are biphase clock flip-flops configured to perform frequency division. They operate at frequences up to 5.77 GHz, which is about the highest division speed yet reported for integrated circuits operating at room temperature. These circuits have been fabricated by electron-beam lithography to produce gate lengths of 0.5 micrometer in the MESFET (metal-semiconductor field-effect transistor) switching transistors. Gallium arsenide devices can be used in very-high-frequency signal processing or as interfaces to more complex chips, including VHSIC (very high speed integrated circuits). (*Note*: This and following two paragraphs were prepared by staff.)

In special tests carried out by several research groups (AT&T Bell Laboratories, IBM, Arizona State University, and Japanese investigators) on gallium arsenide devices, experimental evidence indicates that possibly electrons can travel through a semiconductor without being slowed by collisons, that is, ballistically. As pointed out by Robinson (January 1986), at least in theory, the faster the electrons travel through a transistor, the faster the device can switch on and off. Although, to date, there has been no demonstration of how to take advantage of ballistic transport in a practical transistor, the concept does excite visions of ballistic electrons traveling at nearly the speed of light and providing the basis for transistors that can switch trillions of times per second.

Gallium arsenide and silicon transistors each have their own specific advantages. GaAs transistors switch faster than Si transistors and they also emit near-infrared and visible light, a property of value when both optical and electrical functions are combined in one chip. In many other respects, the GaAs devices are inferior to their silicon counterparts. Researchers have recently found how to effect epitaxial growth of crystalline GaAs layers on silicon wafers and thus combine the properties of both semiconductors, particularly in the manufacture of high-speed microelectronic chips.

The potential of GaAs and other III-V semiconductors are well portrayed by Yablonovitch (see list).

See also **Microelectronics**; **Semiconductor**; and **Transistor**.

Trihalides: The trifluoride and trichloride, AsF_3 and $AsCl_3$, are liquids at room temperature and the tribromide and triiodide, $AsBr_3$ and AsI_3 are solids, although the former melts at 31°C. Like the analogous phosphorus compounds, they have pyramidal structures. Their hydrolysis in aqueous solution is not quite complete, consistent with their greater ionic character (than the phosphorus halides), as is the fact that As^{3+} is precipitated from their solutions as the sulfide. The only stable binary pentahalogen compound of arsenic is the pentafluoride, AsF_5, a colorless gas, which like the trihalides, is less readily hydrolyzed than the corresponding phosphorus compound. A very unstable pentachloride, $AsCl_5$, has been reported. The mixed halide AsF_3Cl_2 can be made by passing chlorine into ice-cold arsenic trifluoride.

Arsenic (III) Oxide: As_4O_6, exists as tetraarsenic hexoxide in the solid state and in the vapor to above 800°C, where dissociation to As_2O_3 commences. It is somewhat soluble in H_2O (about 20 g/l at 25°C), and its solutions have some acidic properties, although the acid has not been isolated and its formula is probably not $As(OH)_3$, the form used for convenience in writing reactions. It is an amphiprotic substance, since, as stated above, As^{3+} is precipitated by H_2S from acid solutions as the sulfide, while the salts, the arsenites (containing the ion AsO_3^{3-}), are readily formed. Their solubility in H_2O varies across the periodic table, those of the alkali metals being very soluble, those of the alkaline earth metals less so, and those of the heavy metals essentially insoluble. Arsenite ion probably exists as $As(OH)_4^-$ in solution.

Arsenic (V) Oxide: As_4O_{10} is a white solid, decomposes at 315°C, isomorphous with phosphorus pentoxide, P_2O_5, but not produced by a simple oxidation of As_2O_3. It is made by dehydration of arsenic acid or $As_4O_{10}\cdot 4H_2O$. It hydrates to give arsenic acid, $H_3AsO_4\cdot\frac{1}{2}H_2O$. This acid is only slightly weaker than phosphoric acid, which it resembles in forming a wide variety of polyacids. It also forms primary, secondary, and tertiary (ortho) arsenates. Raman spectral studies of concentrated arsenic acid solutions in H_2O have a strong band assigned to the —OH group, whence it is inferred that the acid is present in different forms in concentrated and dilute solutions. Many arsenates are converted by ignition into pyroarsenates, e.g., calcium pyroarsenate, $Ca_2As_2O_7$, and metaarsenates are also known.

Direct fusion of the elements yields a number of arsenic sulfides, including As_4S_3, As_4S_4, As_2S_3, and As_2S_5, the last two being obtained also by precipitation from arsenic(III) and arsenic(V) solutions, respectively. The trisulfide dissolves in alkali sulfide solutions to form thioarsenites:

$$As_2S_3 + 3S_2^{2-} \rightarrow 2AsS_3^{3-} + S$$

while with polysulfides it forms thioarsenates:

$$As_2S_3 + 2S_2^{2-} + S \rightarrow 2AsS_4^{2-}$$

Organoarsenic Compounds: The largest group of organic arsenic-containing compounds is the arsonic acids $RAsO(OH)_2$, where the R may be alkyl, aryl, or heterocyclic groups and their salts. In addition to specific compounds mentioned under the uses of arsenic, some organoarsenic compounds include methylarsine CH_3AsH_2; methylarsenic tetrachloride CH_3AsCl_4; diphenylarsenic peroxide $(C_6H_5)_2$ AsOOAs $(C_6H_5)_2$; triphenylarsenic dihydroxide

$$(C_6H_5)_3As(OH)_2;$$

dimethylarsine borane $(CH_3)_2AsHBH_3$; and ethoxydichloroarsine $C_2H_5OAsCl_2$.

Uses: Future production of As_2O_3 will be influenced by the ability to handle ores in the manner required to comply with environmental restrictions. As_2O_3 is available in two grades: (1) crude, 95% As_2O_3; and (2) refined arsenic, 99% As_2O_3. Domestic supplies of the United States are supplemented by imports from Sweden, France, and Mexico. Commercial arsenic metal is produced chiefly by the United States and Sweden.

Arsenic trioxide finds major use in the preparation of other compounds, notably those used in agricultural applications. The compounds monosodium methylarsonate, disodium methylarsonate, methane arsenic acid (cacodylic acid) are used for weed control, while arsenic acid, H_3AsO_4, is used as a desiccant for the defoliation of cotton crops. Other compounds once widely used in agriculture are calcium arsenate for control of boll weevils, lead arsenate as a pesticide for fruit crops, and sodium arsenite as a herbicide and for cattle and sheep dip. In some areas, arsenilic acid has been used as a feed additive for swine and poultry. Restrictions on these compounds vary from one country and region to the next.

Refined arsenic trioxide is used both as a fining and decolorizing agent in glass. As_2O_5 and arsenic acid are used in the manufacture of chromated copper arsenate which is used extensively as a wood preservative.

Indium arsenide, gallium arsenide, and gallium arsenide phosphide find use as semiconductors. For these materials, the starting arsenic source must be extremely pure. Arsenic trichloride and arsenic hydride (very high purity) find application in the production of epitaxial gallium arsenide. Also, in various combinations with iodine, germanium, selenium, sulfur, tellurium, and thallium, arsenic will form a group of glasses with very low melting points.

The applications of arsenic as a metal are quite limited. Metallurgically, it is used mainly as an additive. The addition of from $\frac{1}{2}$ to 2% of arsenic improves the sphericity of lead shot. Arsenic in small quantities improves the properties of lead-base bearing alloys for high-temperature operation. Improvements in hardness of lead-base battery grid metal and cable-sheathing alloys can be obtained by slight additions of arsenic. Very small additions (0.02–0.05%) of arsenic to brass reduce dezincification.

Toxicity: Although metallic arsenic and arsenic trisulfide may be handled, as in the case of most arsenical compounds, skin contact should be avoided. Arsine requires extreme caution in handling because of its very high toxicity. In handling arsenic and its compounds, reducing conditions should be avoided because these may give rise to the undesired formation of arsine. Wherever arsenic and its compounds are present as dusts or vapors, proper ventilation and respirators are mandatory. Good housekeeping and appropriate hygienic practices should also be observed.

Arsenic is commonly found in small amounts in the tissues of plants and animals. A human body may contain as much as 20 mg (As_2O_3). No role in natural biological phenomena has been found for As. Although the element may be present in seawater to the extent of 0.006–0.03 ppm, it may be ten times as high in estuaries. Shellfish tend to accumulate the arsenic from the large amount of seawater with which they come in contact. Oysters may contain 3–10 ppm. However, shellfish of the same species grown in different localities show wide variations in arsenic content, suggesting that it is an accidental constituent which the organisms learn to tolerate. Its lack of function in the human body is suggested by the fact that it tends to accumulate in the hair and nails which are essentially nonliving.

As and Gene Amplification: Research by Te-Chang Lee and associated scientists have reported that arsenic acts specifically in the progression phase of carcinogenicity. Findings indicate that arsenic may be related to its ability to cause gene amplification, but not gene mutations. Data collected regarding humans occupationally exposed to arsenic indicates that exposure to As appears to act at a late stage in the carcinogenic process. Thus, the scientists postulate that amplification of an altered or activated oncogene may be a late stage in neoplastic progression. The hypothesis would explain why arsenic is not an effective complete carcinogen, initiator, or tumor promotor. There may be other substances, in addition to As, that act in this manner. (See Lee reference listed.)

Additional Reading

Carapella, S. C., Jr.: "Properties of Pure Arsenic," in "Metals Handbook," ASM International, Materials Park, Ohio. (Published periodically.)

Lee, Te-Chang, et al.: "Induction of Gene Amplification by Arsenic," *Science*, 79 (October 20, 1989).

Perry, R. H., and D. Green: "Perry's Chemical Engineers' Handbook," 6th Edition, McGraw-Hill, New York, 1988.

Staff: "Handbook of Chemistry and Physics," 73rd Edition, CRC Press, Boca Raton, Florida, 1992–1993.

Yablonovitch, E.: "The Chemistry of Solid-State Electronics," *Science*, 347 (October 20, 1989).

ARSENOPYRITE. The mineral arsenopyrite is a sulfarsenide of iron corresponding to the formula FeAsS. A variety in which some of the iron is replaced by cobalt is known as danaite. It crystallizes in the monoclinic system but twinning produces pseudo-orthorhombic crystals. Its hardness is 5.5–6; sp gr, 6.07; luster, metallic color, silvery-white to steel-gray, but usually with a yellow to gray tarnish; streak, black. Arsenopyrite is a common mineral with tin and lead ores and in pegmatites, probably having been deposited by action of both vapors and hydrothermal solutions. It is a widespread mineral, well-known deposits occurring in Austria, Saxony, Switzerland, Sweden, Norway; Cornwall and Devonshire, England; Bolivia; in the United States at Roxbury, Connecticut; Franklin, New Jersey; Paris, Maine; Emery, Montana; and Leadville, Colorado. Danaite was first found in Franconia, New Hampshire, by J. D. Dana, for whom it was later named. Arsenopyrite also is known as *mispickel*, an old German term whose exact derivation is unknown.

ARTEMISIA. Genus of the family *Carduaceae* (aster or thistle family). Many of the nearly 300 species of *Artemisia* have been cultivated. Southernwood, *Artemisia abrotanum*, is cultivated in gardens for its delicate foliage and aromatic odor. Another and a homely species, *Artemisia vulgaris* or mugwort, is also frequently cultivated, as is *Artemisia absinthium*, a native European perennial plant. All contain volatile oils. That from *Artemisia absinthium*, oil of wormwood, is a powerful drug, capable of causing violent convulsions when taken even in small doses. It is used to flavor the alcoholic beverage absinthe, a liquor capable of producing much the same effects as the drug. The sage-brushes of the western United States are all species of *Artemisia*; like other species they contain an abundance of aromatic oil. From *Artemisia dracunculus*, tarragon is obtained. This is used as a condiment, and for flavoring vinegar and mustard.

ARTERIES AND VEINS (Vascular System). Malfunctioning blood vessels predispose several of the major fatal and disabling diseases. These include *cardiovascular diseases*, as typified by ischemic heart disease—angina pectoris and acute myocardial infarction; and *cerebrovascular diseases*, as represented by cerebral thrombosis and cerebral hemorrhage. A general description of the blood circulatory system is given in the entry on **Heart and Circulatory System (Human).**

Blood vessels fail, fully or partially, in several ways. Particularly in connection with the arteries, a process known as *atherogenesis* may occur. This is a slow process and progresses with the age of the individual. Over a period of time, plaquelike lesions form in the arterial walls, reducing both the elasticity and useful diameter of the vessel. This, in turn, diminishes the amount of blood that can be conveyed by a given artery to the organ which it is supplying and also affects the manner in which the artery must respond in concert with the actions of the heart. Arteries so damaged are more prone to obstructing the flow of blood when thrombi or emboli are present, causing heart damage that may terminate in death (heart attack); or causing brain damage that also may terminate in death or in severe paralysis and mental degradation (stroke). Blood vessels also may leak or rupture, particularly if aneurysm is present, causing damage to surrounding tissue.

In describing vascular disorders, certain terms are frequently used. These include:

Thrombus—a blood clot formed within the heart, or in a blood vessel, remaining at its site of origin. Once a thrombus migrates by way of the bloodstream to a different site and when two or more thrombi agglomerate, the term *embolism* may be used. Thrombi and emboli also are sometimes called occlusions—as in coronary occlusion.

Spasm—coronary artery spasms are described in entry on **Ischemic Heart Disease.**

Embolism—the obstruction of a blood vessel by an embolus, often a migrated blood clot or other debris. If not absorbed, once detached from its place of origin, an embolus may circulate freely in the bloodstream until it finally reaches a vessel through which it cannot pass. An embolus on the arterial side of the heart, such as may originate in chronic valvular disease, will lodge in one of the systemic arteries of the body. On the venous side, an embolus, usually arising in some area of infection, will pass into the heart and thence to the arteries of the lungs. In either situation, the consequences are similar—the blood supply to the region is cut off and, unless adequate alternative channels are available, death or grave debility will result. In the case of the large arteries to such organs as kidneys, brain, and heart, there is only limited possibility of collateral circulation from the outset, and infarction results. An infarct is defined as an area of necrosis in a tissue caused by obstruction of circulation to the affected area. Myocardial infarction is defined as the death of an area of tissue in the heart muscle, caused by decreased blood supply to the heart. In the case of emboli in other parts of the body, unless treated, the final consequence may be gangrene.

Embolism may be caused by means other than blood clotting. Fat embolism is seen in severe bone injury, where fat globules released from the pulped bone marrow may occlude arteries in the brain and/or lungs. Air embolism may form when air is accidentally aspirated into the veins, especially those of the neck. Fragments of tumors may be carried as emboli in the circulation to distant parts. These may be too small to cause immediate effects of emboli, but instead they may form the nucleus for secondary growths at some distance from the primary tumor.

Aneurysm—a sac or pouch filled with blood which protrudes from the wall of an artery, a vein, or the heart. In a *true* aneurysm, the wall of the sac consists of at least one of the layers of tissue that make up the wall of the blood vessel. In a *false* aneurysm, all layers of the artery have ruptured, but blood is retained in connective tissue surrounding the artery. See also **Aneurysm.**

Hemorrhage—the escape of blood from a ruptured vessel. This may vary from seepage to a massive flow. See **Hemorrhage.**

Arteries

The arteries are complex, composite tubes that carry blood from the heart to the capillaries. In the adult, the largest of the arteries, the aorta and pulmonary artery, are indeed quite large—up to 1.2 inch (3 centimeters) in diameter. Likening an artery to a pipe or hose is a gross oversimplification because an artery is a dynamic structure that incorporates functions that go far beyond the simple task of a conduit. For example, the elastic properties of an artery, unlike a rigid pipe, make it possible for the artery to accommodate large and cyclic changes in blood flow that are characteristic of normal heart pumping action. With the capability of continuously changing the orifice (channel for transporting the blood), the artery assists in the outward distribution of blood from the heart through what might be called a squeezing action. Further, the action of the artery varies the resistance to flow of blood and thus the back pressure on the heart; the pliable nature of the artery resists rupturing. The muscular tissue of the artery is generously supplied with nerve fibers (*vasomotor nerves*), which in essence are a control over the relative or effective pliability of the artery.

The normal, healthy artery is composed of three coats, each of which is made up of additional layers and thus is quite unlike and much more complex than what one may envision when thinking of a pipe or tube. (1) The *inner coat (tunica intima)* consists of three layers—(a) endothelial cells: (b) connective tissue, which occurs only in the larger arteries; and (c) elastic fibers with microscopic perforations (*fenestrated membrane*). (2) The *middle coat (tunica media)* consists of muscular and elastic tissue. This middle coat provides the artery with extensible and elastic properties. These characteristics enable the artery to expand upon receiving blood at each contraction of the heart and, in turn, to squeeze the blood forward as the orifice is adjusted to receive more blood at another contraction of the heart. (3) An *external coat (tunica externa* or *adventita)* consists of areolar connective tissue made up of smooth muscle cells, either scattered or in the form of longitudinally arranged bundles.

When empty, the arteries do not collapse; the orifice for carrying the blood remains open. However, when an artery is cut, as from an injury, the muscular coat does contract to some extent, thus narrowing the opening of the wound and permitting a clot to form to close (plug) the orifice, an action that is mandatory to the halting of hemorrhage. In the body, a majority of the arteries are accompanied by one or more veins which are contained within a protective sheath of connective material for providing support to these vessels. Some authorities classify arteries as (1) *elastic arteries*, which are the larger vessels that carry blood from the heart to arteries of lesser diameter. The property of elasticity is a prominent feature of these vessels; and (2) *muscular arteries*, which are of smaller diameter that distribute blood to specific organs. The muscular characteristics of these vessels provide ability to reduce (by contraction) or expand (by relaxation) the blood flow in accordance with the specific demands of the organs being supplied. Arteries also have their own blood supply because of energy requirements. The arteries, capillaries, and veins so involved are identified as *vasa vasorum*. As with other blood vessels, the arteries divide and subdivide in a complex branching system. The smallest of these branches are called *arterioles*, and it is at the distal ends of these tiny arterioles that the capillaries begin. As the arteries become smaller, the proportion of smooth muscle to elastic tissue increases.

Arteriosclerosis and Atherosclerosis

Commonly referred to as *hardening of the arteries*, this is a condition that exists when the walls of the blood vessels thicken and become infiltrated with excessive amounts of mineral and fatty materials. Arteriosclerosis occurs in all races and among people living in all climates. The disorder tends to progress with the age of the patient and is sometimes considered a disease of the middle-aged and elderly, but it should be quickly observed that coronary artery disease is responsible for nearly one-third of the deaths of persons between the ages of 35 and 65. The disorder has been known for centuries, although it was not even partially understood until the last half of the 1800s, when the first hypotheses as to its etiology were proposed. The complex and still only partially understood process which causes arteriosclerosis is called *atherogenesis*, particularly as the process applies to the major arteries, such as the coronary arteries. "Hardening" of the major arteries is commonly referred to as *atherosclerosis*. Because of the great importance of this disease as a predisposing condition for heart attack and stroke, intensive research into the etiology, diagnosis, treatment, and prevention of atherosclerosis is being conducted worldwide. Although considerable progress has been made in recent years, the etiology of the disease has not been fully explained and, in fact, there are several viewpoints on this topic. See also **Ischemic Heart Disease;** and **Cerebrovascular Diseases.**

Familial Hypercholesterolemia. In terms of the biological knowledge of the arteries, there is one exception to the foregoing statement, namely, the discovery of the genetic mechanism that is responsible for familial hypercholesterolemia (FH). For their discovery of the LDL (low density lipoprotein) concept and the biological pathways involved in FH, Michael S. Brown and Joseph L. Goldstein (University of Texas) received the Nobel Prize in Physiology or Medicine in 1985. The work of these researchers commenced in 1972. As acknowledged by Brown and Goldstein in their acceptance address, their work was buttressed by the findings of earlier investigators.[1] Rather than directing their efforts to a general relationship between cholesterol and cardiac disease, as has been the case of numerous statistical studies over recent years, Brown and Goldstein concentrated their attention on FH, a disease in which the concentration of cholesterol in the blood is elevated many times above normal and heart attacks occur early in life. The researchers postulated that this dominantly inherited disease was the result of a failure of end-product repression of cholesterol synthesis. Brown and Goldstein used the techniques of cell culture to explain the postulated regulatory defect in FH. Such investigations led to the discovery of a cell surface receptor for a plasma cholesterol transport protein (LDL) and to the explanation of the mechanism by which the receptor mediates feedback control of cholesterol synthesis. The gene encoding for the LDL receptor was shown to be defective in the case of individuals with FH. These defects were shown to disrupt the normal control of cholesterol metabolism. Further study of the LDL receptor led to an understanding of receptor-mediated *endocytosis* (process by which cells communicate with each other through internalization of regulatory and nutritional molecules). This process differs from previously described biochemical pathways because it depends on the continuous and highly controlled movement of membrane-embedded proteins from one cell organelle to another (a process now termed *receptor recycling*). It was learned that many of the mutations in the LDL receptor (of the type occurring in FH patients) disrupt the movement of the receptor between organelles. Brown and Goldstein concluded that these mutations define a new type of cellular defect that has broad implications for normal and deranged human physiology.[2]

Familial hypercholesterolemia was first described by Carl Müller (1938) who referred to its cause as an "inborn error of metabolism." Two forms of the disease (heterozygous and homozygous) were first noted by Khachadurian and by Fredrickson and Levy. The heterozygous form is the less severe of the two types and is present in persons who carry a single copy of a mutant LDL-receptor gene, numbering one person in 500 persons worldwide. Such individuals have a twofold increase in the number of LDL particles in plasma at birth and heart attacks become commonplace at ages 30 to 40 years. These individuals are 25 times more prone to suffer myocardial infarctions by age 60 than are members of the general population. Fewer persons (about 1 in 1 million) inherit two mutant genes at the LDL receptor locus, one from each parent. Needless to say, this form of the disease is much more serious than the heterozygous form. At birth, these persons have 6 to 10 times the normal concentrations of plasma LDL at birth and frequently suffer heart attacks during childhood. These findings in connection with FH indicate proof that high levels of plasma cholesterol can produce atherosclerosis in humans. However, the cause of heart attacks among the general populace, as related to cholesterol levels, still lacks the biological explanation and proof that some experts would like to see. As of the late 1980s, such proofs remain to be found, although there is voluminous statistical information to support the cholesterol connection. This is covered later in this article.

Traditional Concepts of Atherogenesis. In the late 1980s, the German physiologist, Rudolf Virchow, suggested that atherogenesis is a process involving the deposition of cholesterol and other lipids in the arterial wall, resulting from an abnormally high concentration of lipids in the blood. Virchow further suggested that endothelial cell injury initiated atherogensis. During the intervening years, much clinical and experimental evidence has been collected to support the hypothesis and, in fact, during the first half of the present century, Virchow's hypothesis served as the basis for preventive cardiology. In more recent years, several other hypotheses have been proposed, but as of the late 1980s a fully satisfactory resolution of these various scientific viewpoints has not occurred.

From research on laboratory animals and human autopsies, much has been learned concerning the arterial deposits (plaques) and the structural and compositional changes which have occurred when diseased arteries are examined. Since atherogenesis is a slow process and non-invasive instrumental methods for observing living tissues over a long period are limited, a fully satisfactory step-by-step description of the total process, including the initiating event or events, remains to be delineated.

Multiple-Risk or Risk-Factor Hypothesis

In the early 1980s, essentially by consensus among many scientists rather than by a bank of convincing scientific evidence and in lieu of a better understanding of the etiology of atherosclerosis, a multifaceted approach to the disease was considered the most practical guideline for professionals in the field. This hypothesis continues as the most influential concept in preventive cardiology. The risk factors include:

1. High serum concentration of blood lipids (hyperlipidemia), including the role of cholesterol.
2. Hypertension.
3. Cigarette smoking.
4. Insufficient exercise.
5. Certain genetic disorders.
6. Excessive intake of animal proteins, among others.

All of the listed risk factors appear to be predisposing factors for arteriosclerosis in susceptible individuals.

Hyperlipidemia—Role of Cholesterol. Numerous studies (see listed references) have provided statistical evidence to demonstrate the correlation of high levels of cholesterol with high levels of serum cholesterol. The linkage to cholesterol, in addition to the previously reported studies of Brown and Goldstein, is largely based upon the observation that atherosclerotic plaques contain large deposits of cholesterol and upon data from epidemiological studies.

Cholesterol is a steroid bearing an alcohol group. It is unstable in water and essentially all of it is carried in the blood as lipoproteins. Some cholesterol is required by the body as a precursor for the manu-

---

[1] In 1901, after studying a patient with black urine, Garrod, a physician, suggested that a simple mutant gene can produce a discrete block in a biochemical pathway, which Garrod called an "inborn error of metabolism." Garrod's exceptional insight preceded by some forty years the "one-gene-one-enzyme" concept of Beadle and Tatum. Later the Nobel Prize winning Linus Pauling and a physician, Vernan Ingram, through studies of patients with sickle cell anemia, showed that mutant genes alter the amino acid sequences of proteins.

[2] As an aside, it is interesting to note that since it was first isolated from gallstones (1784), cholesterol has been of fascinating intrest to scientists. Thirteen Nobel Prizes have been awarded to scientists who devoted most of their careers to studying cholesterol.

facture of certain steroids, such as bile acids and several hormones. Many investigators insist that the amount of cholesterol produced is governed in some way by the dietary intake of cholesterol and there is much statistical information to support this view.

Lipoproteins are large, complex polar molecules. Carried on the inside are the nonpolar molecules, such as the triglycerides (esters of glycerol and three long-chain fatty acids) and esters of cholesterol. It has been estimated that much of the cholesterol is carried in ester form. The remainder exists as nonesterified cholesterol contained on the surface of the lipoproteins. Currently, the lipoproteins are categorized as:

LDL (low-density lipoprotein or beta fraction)—rich in cholesterol, but poor in triglycerides. Identified by Brown and Goldstein in connection with HF disease, as previously mentioned.

VLDL (very-low density lipoprotein or prebeta fraction)—rich in triglycerides and contains substantial cholesterol.

HDL (high-density lipoprotein or alpha fraction)—normally contains 20–25% of the total cholesterol, but appears to exert a protective effect. It has been suggested that HDL inhibits cholesterol uptake by smooth muscle cells, or by transporting cholesterol out of smooth muscle cells to its site of metabolism in the liver.

The aforementioned ratio of total cholesterol to HDL cholesterol has been found to be a useful predictor of coronary artery disease and a more reliable indicator than the level of either fraction alone. A favorable ratio is 4.5 : 1 or lower.

Investigators have found that the LDLs carry much of the cholesterol found in the blood and play a major role in cholesterol metabolism (in cells other than the liver) and participate in the formation of atherosclerotic lesions. The function of the HDLs appears to be that of transporting cholesterol from the peripheral tissues to the liver. Although the HDLs may be incorporated into LDLs or VLDLs and recycled to the peripheral tissues, some investigators believe that the HDLs may provide a means for removal of cholesterol from the tissues and thus decrease the probability of its terminal deposition as atherosclerotic plaques.

It has been found that specific receptors for the HDLs are contained in human fibroblasts and certain other cells. The interaction of the HDLs with receptors is considered by some researchers as an initial step prior to the degradation of cholesterol and thus suppress cholesterol synthesis in these cells.

As research continues, the explanations of the lipoproteins multiply, but currently with no common denominator within view. It is apparent that once a key is found, this will be the foundation of preventive measures. Although it is possible that such a key may be diametrically opposed to current preventive procedures, most practitioners are staying with the treatment of hyperlipidemia. Hyperlipidemia is treated primarily by dietary restriction, an approach that seems effective when elevations of both cholesterol and triglyceride levels are secondary to obesity and do not reflect a primary genetically determined form of hyperlipidemia. It has been found that weight reduction, achieved by restriction of food calories and alcohol intake, is the principal cause of the decrease in serum lipids in patients who respond to such therapy. Serum cholesterol can be reduced modestly by diets that are low in cholesterol or that substitute polyunsaturated fats (largely of vegetable origin) for saturated fats (largely of animal origin).

One of the most recent findings, described shortly, is the apparent relationship between fish consumption and coronary heart disease.

It has been a general practice of many cardiologists to treat high serum cholesterol levels either by dietary control, by drugs, or by a combination of the two. Cholestyramine is a high-molecular-weight resin that binds bile acids in the gastrointestinal tract and has been used quite extensively even though in some patients it has discomforting side-effects and is relatively costly.

A suggested regimen for dieting followed by many cardiologists is:

1. Avoid overweight by balancing caloric intake with exercise-mediated energy expenditure.
2. Reduce overall fat intake from about 40% to about 30% of total calories.
3. Reduce intake of saturated fat so that saturated fat and unsaturated fat each accounts for about 10% of caloric intake.
4. Reduce cholesterol intake to about 300 mg per day.
5. Reduce consumption of refined and processed sugars by about 45% so that such sugars account for about 10% of total calories; increase consumption of complex and unrefined carbohydrates from about 28% to about 48% of caloric intake.
6. Limit sodium intake to about 5 grams per day.

This diet, of course, is not exclusively directed to reducing cholesterol levels, but to accommodate other heart disease risks as well.

Hypertension as a Risk Factor. The risk of hypertension in coronary disease probably rises from the increasing physical stress placed on the arterial wall, thus enhancing the process of arterial wall injury. One hypothesis stresses how factors such as variations in intra-arterial pressure and uneven flow patterns account for the distribution of atherosclerotic plaques. A difference in arterial pressure and its effect on plaque formation is demonstrated by the fact that high-pressure systemic arteries display much more damage than that found in the low-pressure pulmonary arteries. See entry on **Hypertension.**

Smoking as a Risk Factor. It has been established that smoking enhances platelet stickiness and causes coronary vasoconstriction produced by nicotine and hypoxemia induced by inhalation of carbon monoxide. In contrast with the risk of smoking (cumulative) in lung cancer, the risk in coronary disease appears to be reversible upon discontinuation of smoking.

Lack of Exercise as a Risk Factor. The popularity and broad acceptance of the importance of exercise in preventing coronary artery disease greatly exceeds the scientific evidence. For example, there is no evidence that coronary collateral vessels develop or that atheromatous changes regress as a result of exercise. Exercise as a factor in weight reduction can lead to a reduction in serum lipid levels, but as previously described, lipid levels remain a topic of controversy. **It should be noted here that the importance of exercise for the rehabilitation of patients who have had a clinical coronary disease, recently or remotely, had different goals and techniques than a program designed for a healthy person as a preventive measure.**

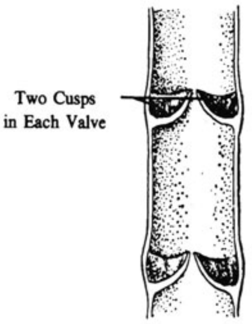
Although stress, personality, and behavior patterns as causative factors in coronary artery disease have received wide public recognition and recognition by some physicians as well, relatively little attention has been given to this topic by pathologists and epidemiologists, largely because these factors are difficult to study.

Response-to-Injury Hypothesis. The role of injury to endothelial cells, part of Virchow's original hypothesis on atherogenesis, has been revived in recent years and current findings are based largely upon studies of lesions of animal arteries. The present version states that plaques form in response to frequently recurring injuries to arterial walls. It has been noted that nearly any type of chronic injury to animal arteries will progress to the formation of lesions that are much like the plaques found in human arteriosclerosis. This hypothesis suggests that injury to the endothelium, causing desquamation (shedding of superficial cells), is the primary event. Smooth muscle cells then migrate from the media into the intima through fenestrae in the internal elastic lamina and there undergo active proliferation within the intima. The adherence of platelets to exposed connective tissue may cause the formation of platelet aggregates, or microthrombi. As the lesion progresses, fibrosis, lipid deposition, and calcification may ensue, thus yielding the type of complex plaque noted in diseased arteries.

Based largely upon this hypothesis, platelet-inhibiting drugs, such as aspirin, are in use.

A Genetic Approach to Heart Disease

The risk-factor hypothesis previously described is in a way an "after the fact" approach to minimizing heart disease, requiring ambitious public awareness programs to alert the population of the risks (high blood pressure, high levels of serum cholesterol, smoking, etc.) and thus by altering the habits of society, ultimately reduce the risk of heart disease. In another approach, some authorities believe that rather than targeting programs that intervene with societal habits (and thus meet with very slow acceptance), the principal effort should be directed toward determining well in advance those persons who are most likely to be naturally prone to heart disease. One very obvious way to do this is to turn the problem over to molecular biology. At a conference of the American Heart Association (Sarasota, Florida, January 1986), an approach taken by California Biotechnology, Inc. (Mountain View, California) was described. This group is seeking genetic markers near the sites of genes that are known to be involved with lipid metabolism and with blood pressure regulation. The target, of course, is to develop highly accurate predictive tests for identifying persons at risk. One spokesman for the group expressed that a test of 90% accuracy for predicting persons who will get cardiovascular diseases will be relatively easy to attain and that a level of nearly 100% accuracy may be possible. Some genetic proof for this approach rests in the fact that there are numerous examples of where heart disease has a familial connection. Specialists in the field predict that genetic markers for heart disease may become as common and as well accepted as those which currently apply to Huntington's disease and muscular dystrophy, among other well known diseases with a hereditary nature.



Large vein, showing how valves composed of two cusps occur along the vein and prevent the backflow of blood in these vessels.

Buerger's Disease

This is an arterial disease (*thromboangitis obliterans*) which also may involve the veins. This is an inflammatory disease which is followed by obliteration of the lamina. This results in impairment of the circulation to the extremities, manifested by coldness, cyanosis, pain, and, if untreated and uncontrolled, eventually gangrene of the affected part. Treatment is directed toward improving the circulation of the extremities by special exercises, elevation of the affected part, drug therapy, and occasionally surgical measures which remove vessel-constricting impulses by interfering with their nervous pathways.

Aortic Aneurysms—Synthetic Grafts

The presence of aneurysms in any segment of the ascending aorta often occur, resulting from prior aortic dissection, atherosclerosis, and uncommonly aortitis, syphilis, and trauma. The ascending aorta is also subject to diffuse fusiform dilatation, which is associated with aortic valvular regurgitation, a condition called *annuloaortic ectasia*. Effective treatment often involves synthetic graft replacement of the ascending aorta, sometimes also requiring replacement of the aortic valve. Atherosclerotic aneurysms of the descending thoracic aorta are usually asymptomatic and have little tendency to rupture or to cause pressure symptoms. They are seldom removed unless large (greater than 10 cm), or progressively enlarging and causing symptoms.

Aneurysms are found more often in the abdominal aorta, usually an atherosclerotic lesion most frequently found in men over 60 years old. Noninvasive techniques are reliable for detecting these aneurysms. Although not usually required, angiography is sometimes used prior to surgery. The usual procedure is resection of the aneurysm and replacement with a prosthetic graft. Mortality of the procedure ranges from 2 to 5%, death usually caused by myocardial infarction. Patients considered of possible high risk are evaluated thoroughly prior to the final decision for surgery. Such clinical evaluation may include treadmill exercise testing, radionuclide scanning, and coronary arteriography. In the case of a ruptured aneurysm, emergency surgery is required, which involves an operative mortality risk of 25 to 50%.

The occurrence and treatment of acute peripheral arterial occlusive disease and chronic occlusive arterial disease is beyond the scope of this encyclopedia.

Other diseases of the arteries include *aortitis*, which is an inflammatory condition sometimes defined as a nonspecific form of pathologic change that represents a final common effect of many disorders arising from such underlying causes as syphilis and bacterial infection. Aortitis also may be associated with rheumatoid arthritis, lupus erythematosus, scleroderma, psoriasis, ulcerative colitis, and Crohn's disease. Important forms of primary aortitis are the ideopathic diseases known as *Takayasu's aortitis* and *giant cell arteritis*. Takayasu's aortitis (also known as pulseless disease) principally affects young women and is mainly seen in Asia. Giant cell arteritis is characterized by an accumulation of multinucleated giant cells in the arterial walls. Usually found in the elderly, giant cell arteritis can be treated effectively with corticosteroids.

Venous Disorders

Varicose Veins. These are veins that have lost their elasticity and, as a consequence, are irregularly enlarged and swollen. They have a dilated, lumpy, twisted and tortuous appearance. The overlying skin may be affected with ulcers. Varicose veins are most often seen in the legs of middle-aged and older persons, although certain conditions, such as pregnancy, may cause them to appear in younger adults. The dilation of the veins results from the inability of the weakened venous walls to withstand the pressure of the blood within the veins.

If the veins were simply continuous tubes running from the legs to the heart, the weight of a column of blood carried this high would press out on the leg veins when the individual stood erect. Normally, the column of blood is broken by the presence of valves which prevent the full weight of the blood from causing undue pressure on the veins in the leg. See accompanying diagram. If a vein loses its elasticity, it will become distended, and the valves will fail to close completely. The weight of the blood in that vein then presses out on the walls of the vein, causing even more distention; and a reversal of flow of blood in that vein may occur. The veins also may become swollen when a venous constriction prevents the normal emptying of the blood.

The classical sign of varicose veins is the actual appearance in the legs of swollen, tortuous, blue veins. The enlargement may affect a short segment of a single vein, or nearly all of the veins in the entire leg. When a systemic disease is responsible for the disorder, it usually appears in both legs to an approximately equal degree. *Phlebitis* (inflammation of a vein), constriction, injury, or obstruction in the veins in one leg, however, will affect only the one leg. Aside from disfiguration, varicose veins cause appreciable physical discomfort. Both dull and stabbing pains may be noted, and the entire limb may become quite swollen. When the condition has existed for some time, the veins sometimes become toughened and thick, so that they feel firm to the touch. More often, however, they are soft and elastic, except at the hard knotty swellings which occur in the regions of the valves. Ulceration and bleeding may leave large black and blue areas beneath the skin. In men, a type of varicosity may occur in which the vein's in the scrotum are affected. In this condition, known as *varicocele*, the scrotum contains a soft, tumor-like mass of swollen venous materials. Similarly, varicosity of the veins in the rectum is known as *hemorrhoids* (piles).

In severe cases of varicose veins, it may be necessary to surgically remove portions of a vein that are particularly bothersome, or strip out the entire varicosed vein. In milder cases, merely tying off (ligate) the varicose veins to relieve pressure may suffice. Occasionally, a combination of an operative procedure and sclerosing solutions which close off small veins may be required.

Raynaud's Disease. A vascular disorder characterized by paroxysmal cyanosis of the fingers in response to cold or to emotional stimuli. Toes, ears, chin, may also be affected. The disease appears most commonly in women in the third decade of life, and familial connections have been described. The cyanosis is due to interference with the blood supply through constriction of the vessels (vasospasm) resulting from abnormalities of the nervous control of the vessel walls.

Venous Thrombosis. Two forms of venous thrombosis are presented: (1) Lesions in *superficial veins* may be caused by tightly adherent thrombi. There may be significant inflammation and in some cases cellulitis, but these thrombi rarely penetrate to deep veins to cause pulmonary emboli. Treatment includes application of local heat and possibly administration of antibiotics where infection is suspected. (2) *Deep venous thrombi* may lead to pulmonary embolism, the latter being a life-threatening situation. Some physicians describe deep venous thrombi as "silent" because they may not be detected by routine clinical examination. In fact, the primary site may be difficult to fix even after the diagnosis of pulmonary embolism. However, when present, clinical signs include distended veins, localized areas of tenderness over venous structures and, in the calf, notation of increased heat, edema, and sometimes discoloration. Some conditions mimic the symptoms of deep venous thrombi, including rupture of calf muscle fibers or of synovial cysts involving the knee. Procedures for diagnosis of deep venous thrombi include venography which will reveal any filling defects. This procedure is used with discretion because the invasion necessary for venography can in itself induce venous thrombosis. Although not applicable to the calf, noninvasive diagnostic measures include impedance plethysmography and ultrasonography. Radionuclide scanning is useful in connection with the calves, but not in the thighs. The treatment of deep vein thrombosis is the prompt administration of intravenous heparin. This drug acts rapidly in preventing further thrombus formation and prevents the release of serotonin and thromboxane A_2 from platelets adherent to those thrombi that embolize to the lungs. See **Anticoagulants.**

Pulmonary Embolism. The formation of emboli is the major consequence that is feared from deep venous thrombi. Although pulmonary emboli are sometimes overdiagnosed, statistics also show that in a high percentage of fatal cases, the disease has gone on unsuspected for quite a period of time. The usual symptoms are apprehension, weakness, faintness or syncope, dyspnea, and a substernal pain that is not distinguishable from that of myocardial infarction. Lung examination may indicate wheezes. To assist in diagnosis, the physician will order blood gases analysis and frequently an emergency perfusion lung scan. However, a number of conditions also alter pulmonary blood flow and such scans are not fully definitive. In recent years, pulmonary arteriography has become the reference standard for the diagnosis of pulmonary embolism. However, the procedure is undertaken with discretion because of risks associated with it. Other diagnostic tools include chest x-ray, electrocardiography, and serum enzyme levels. Treatment of pulmonary embolism includes drugs, such as morphine sulfate or meperidine, to relieve pain and the institution of intravenous heparin therapy. After several days, the patient may be switched to oral anticoagulants, a treatment which in some cases may extend over a year.

Massive Pulmonary Embolism. In this case, large emboli occlude the proximal pulmonary arterial circulation, producing symptoms much like those of myocardial infarction. Usually, electrocardiography will distinguish between the two conditions. Massive pulmonary embolism causes many deaths and, in any case, must be considered a life-threatening emergency. Upon diagnosis, large doses of intravenous heparin are given, along with morphine and oxygen. If there is persistence of shock for over 30 minutes, an immediate pulmonary arteriogram should be done. However, this requires the equipment and skills of an open-heart surgery facility. In severe instances and where massive pulmonary emboli are confirmed, surgery to remove the embolism may be elected to save the life of the patient.

In the management of massive pulmonary embolism and severe deep venous thrombosis, thrombolytic therapy has been relatively recently available. This involves the use streptokinase and urokinase, agents which hasten the dissolution of thrombus. Thrombolytic agents cannot be given concurrently with anticoagulants and they do have a number of side effects. Several years of additional experience will probably be required to assess the values and risks of thrombolytic agents.

See also list of related entries at end of entry on **Heart and Circulatory System (Human).**

Additional Reading

Bashore, T. M., and C. J. Davidson, Editors: "Percutaneous Balloon Valvuloplasty and Related Techniques," Williams and Wilkins, Baltimore, Maryland, 1991.

Braunwald, E., Editor: "Heart Disease: A Textbook of Cardiovascular Medicine," W. B. Saunders, Philadelphia, Pennsylvania, 1992.

Burke, G. L., et al.: "Trends in Serum Cholesterol Levels from 1980 to 1987—The Minnesota Heart Survey," *N. Eng. J. Med.*, 941 (April 4, 1991).

Criqui, M. H., et al.: "Mortality Over a Period of 10 Years in Patients with Peripheral Arterial Disease," *N., Eng. J. Med.*, 381 (February 6, 1992).

Coffman, J. D.: "Intermittent Claudication—Be Conservative," *N. Eng. J. Med.*, 577 (August 22, 1991).

Ernst, C. B., and J. C. Stanley, Editors: "Current Therapy in Vascular Surgery," B. C. Decker, Philadelphia, Pennsylvania, 1991.

Ewy, G. A., and R. Bressler, Editors: "Cardiovascular Drugs and the Management of Heart Disease," Raven, New York, 1992.

Fozzard, H. A., et al., Editors: "The Heart and Cardiovascular System: Scientific Foundations," Raven Press, New York, 1991.

Grundy, S. M.: "Trans Monounsaturated Fatty Acids and Serum Cholesterol Levels," *N. Eng. J. Med.*, 480 (August 16, 1990).

Hurst, J. W., et al.: "The Heart, Arteries and Veins," McGraw-Hill, New York, 1990.

Loscalzo, J.: "Regression of Coronary Athersclerosis," *N. Eng. J. Med.*, 1337 (November 8, 1990).

Mills, J. A.: "Aspirin, the Ageless Remedy?" *N. Eng. J. Med.*, 1303 (October 31, 1991).

Moore, W. S., and S. S. Alm, Editors: "Endovascular Surgery," W. B. Saunders, Philadelphia, Pennsylvania, 1989.

Muller, J. E., and G. H. Tofler: "Circadian Variation and Cardiovascular Disease," *N. Eng. J. Med.*, 11038 (October 3, 1991).

Panza, J. A., Epstein, S. W., and A. A. Quyyumi: "Circadian Variations in Vascular Tone and Its Relation to Alpha-Sympathetic Vasoconstrictor Activity," *N. Eng. J. Med.*, 986 (October 3, 1991).

Sacks, F. M., and W. W. Willett: "More on Chewing the Fat—The Good Fat and the Good Cholesterol," *N. Eng. J. Med.*, 1740 (December 12, 1991).

Wind, G. G., and R. J. Valentine: "Anatomic Exposures in Vascular Surgery," Williams and Wilkins, Baltimore, Maryland, 1991.

Young, J. R. et al., Editors: "Peripheral Vascular Diseases," Mosby Year Book, St. Louis, Missouri, 1991.

Zemel, M. B.: "Altered Platelet Calcium Metabolism as an Early Predictor of Increased Peripheral Vascular Resistance and Preeclampsia," *N. Eng. J. Med.*, 434 (August 16, 1990).

ARTHRITIS (Infectious). An inflammatory joint disease caused by pyogenic (pus-forming) bacteria, mycobacteria, or fungi.

Acute Bacterial Arthritis. The *acute* form of the disease is usually caused by pyogenic bacteria, of which the most common cause is *Neisseria gonorrhoeae*. Of the remaining cases, caused by nongonococcal bacteria, *Staphylococcus aureus, Streptococcus pneumoniae*, and other strep microorganisms are involved. A wide variety of infections which cause bacteremia may predispose acute bacterial arthritis—pneumococcal pneumonia, endocarditis, skin infections, and cholangitis, among others. The condition may develop as the result of joint surgery. As reported by Goldenberg and Reed (see reference), septic arthritis complicating prosthetic-joint surgery has become a serious medical problem since the late 1970s. Early accounts of prosthetic-joint surgery reported a 5 to 8% rate of infection (1972). Later reports indicated a 1 to 4% infection rate. However, a 1984 report indicated that the infection rate may be as high as 30% in surgical revision of prosthetic joints. In most studies of prosthetic-joint infections, the cases have been divided into *early* (occurring in the first postoperative year) and *late* infections. Intraoperative wound contamination is considered a likely source of some early infections, whereas late infections are most likely to be hematogeneously acquired. Staphylococci are the most common pathogen. *Staph, epidermidis* accounts for 40% of prosthetic-joint infections, and *Staph, aureus* for 20%. Multiple organisms and anaerobic bacteria also are often involved. About 12% of cases of acute bacterial arthritis are attributed to gram-negative bacili. An infection of this type is sometimes called *acute septic arthritis*. Related or predisposing factors include urinary or biliary tract infection, neoplastic (tumor-forming) diseases, diabetes mellitus, intravenous heroin abuse, and certain antibiotics and immuno-suppressive drugs. In about 20% of cases, *Salmonellae* are implicated; *Pseudomonas* and *Serratia* species are usually involved in cases of septic arthritis in narcotic addicts. *Haemophilius influenzae* is an uncommon agent in adults, but is a frequent cause of the disease in children. At one time, prior to tight health controls over milk and milk products, brucellosis was a major cause of spinal infectious arthritis.

Bacteria may invade the joint cavities by way of the bloodstream or they may spread from a contiguous focus of osteomyelitis. Purulent discharge may damage articular cartilage, joint capsule, and subchondral bone.

Symptoms of bacterial arthritis are hot, swollen, and painful joints. Onset may be abrupt, particularly in the presence of systemic illness. Fever of a transient or low-grade level may be present. Most frequently involved joints are the knees, followed by hips and shoulders. Symmetry is not usually present, an advantage in terms of diagnosis.

Precise diagnosis usually requires examination and culture of joint fluid. In early phases of the disease, x-rays are of little help. There are several disorders which have similar symptoms and the diagnosis may be complex. Early diagnosis, however, is extremely important if recovery without permanent joint damage is to be achieved. Choice of most effective antimicrobial and antibiotic therapy hinges on accurate diagnosis. Drainage of the affected joint is also important to reduce pressure and to remove cartilage-destroying enzymes present in the pus. Aspiration at intervals over several days may be required. Moderate movement of the joint after the initial phase is encouraged to maintain mobility of the joint. The affected joint should not be subject to weight bearing, however, until inflammation has subsided. With early diagnosis and prompt therapy, full recovery without joint damage can be expected.

Tuberculosis Arthritis. This is uncommon. The disease is caused by the presence of *Mycobacterium tuberculosis*. The joints most often involved are hips and knees in children and vertebral joints in adults. Predisposing factors include diabetes mellitus, lupus erythematosus, narcotic addiction, and intra-articular injections of corticosteroids. Often present for a long period without notice by the patient, the disease can result in serious, permanent joint damage. Diagnosis may require an open synovial biopsy. Therapy includes resting the joint and use of antimicrobial drugs, such as isoniazid and ethambutol.

Fungal Arthritis. This relatively uncommon type of infectious arthritis results from invasion by mycotic agents and varies considerably with geographical location. For example, the disease may be associated with coccidiomycosis in the southwestern United States and with blastomycosis in the southeastern United States. Amphotericin B therapy is usually indicated.

Syphilitic Arthritis. This may occur in congenital, secondary, or tertiary syphilis. It is most commonly seen in infants and very small children with congenital syphilis. In these instances it may lead to destructive osteochondritis. A more moderate form of the disease may develop in older children and adolescents with congenital syphilis. Penicillin therapy is usually indicated.

Lyme Arthritis. This relatively recently recognized disease is attributed to a virus. The disease can result in chronic joint destruction, but in most cases is self-limiting. See also **Lyme Disease.**

ARTHRITIS (Rheumatoid). See **Rheumatoid Arthritis.**

ARTHROPODA. The largest and most diversified division of the animal kingdom, including crustaceans, horseshoe crabs (see **Xiphosura**), insects, scorpions, spiders, centipedes, millipedes (see **Diplopoda**), and other forms.

This phylum is characterized by the following structures: (1) The body is triploblastic and metameric (see **Metamere**), and is further subdivided into regions of which there may be a maximum of three: head, thorax, and abdomen. (2) Supporting structures are developed from the integument and constitute an exoskeleton made up of plates connected by flexible regions for freedom of movement. (3) The appendages are jointed, fundamentally a pair to a segment, which gives the phylum its name. (4) The circulatory system is a combination of tubes and open spaces; the coalescence of the latter to form a haemocoel is accompanied by extreme reduction of the coelom. (5) The eyes are of a form peculiar to the group. (6) The respiratory system consists of air tubes or tracheae, of gills, or of lung books formed of leaf-like expansions of the body wall located in a cavity narrowly open to the exterior. Some species have no special respiratory organs.

The phylum is divided into several classes as follows:

Class *Onychophora*. Soft-bodied wormlike animals. Commonly called *Peripatus*, the name of one genus.
Class *Tardigrada*. The bear animalcules.
Class *Pentastomida*. Parasites known as pentastomids or linguatulids.
Class *Pycnogonida*. The sea-spiders.
Class *Crustacea*. Crabs, lobsters, crayfishes, shrimps, barnacles, woodlice or pillbugs, and other forms.
Class *Xiphosura*. The king crab or horseshoe crab. Also given the name *Palaeostraca*.
Class *Arachnida*. Spiders, mites, ticks, and scorpions.
Class *Diplopoda*. The millipedes.
Class *Pauropoda*. Rare forms allied to the preceding.
Class *Chilopoda*. The centipedes.
Class *Symphyla*. Rare forms allied to the preceding.
Class *Insecta*. The insects. See also **Invertebrate Paleontology.**

A very small proportion of the arthropoda, notably of the insecta, have widespread influence on the health of the human race; either by direct injury but chiefly through the destruction of growing crops and stored food and their capabilities as carriers of disease to humans and domestic animals. In humans, the diseases so produced are typhus and malaria, yellow fever, filariasis, bacillary dysentery, and other enteric infections.

ARTHROSCOPY. A procedure in which a fiberoptic tube, or endoscope, is used to look inside knees, shoulders, or other joints. Arthroscopic surgery entails the use of a microminiaturized video camera and a special instrument in the form of a probe that is about the size of a ballpoint pen and attached to the video camera, approximating the size of a common safety match box. Operations that previously were accomplished through large incisions can be done through several small puncture wounds. For example, the procedure is used frequently in the treatment of carpal tunnel syndrome, a disease that causes pain and numbness of the hand.

ARTICULATION (Communications). In verbal communication, the main purpose of speech is to convey thoughts. In testing speakers (human) and communication systems, a number of tests have been designed for measuring the percentage of words or individual speech sounds uttered by a speaker which are perceived correctly by listeners. For example, the Harvard PB-50 word list employs a set of phonetically balanced words and is widely used. During the testing procedure, the talkers read the word lists over the system under test to a number of listeners. As they hear them, the listeners record the words. Responses are examined and the percentage of words heard correctly is determined. This is termed *percent word articulation*.

During such tests, a number of communication system parameters are explored. Methods have been developed for computing speech intelligibility from system characteristics. The measure of intelligibility computed is termed the *articulation index* (AI). A number of observations concerning speech are taken into consideration by the articulation-index concept: (1) speech must be above threshold of audibility to be perceived; (2) noise that exceeds this threshold masks speech, effectively raising the threshold of audibility; (3) there is an upper limit to the sound pressure which the ear can utilize for perception of speech; (4) frequencies from 200 to 6100 Hz are needed for substantially perfect intelligibility; (5) speech has a 30-decibel dynamic range; and (6) different frequency regions contribute unequally to intelligibility, but the frequency range from 200 to 6100 Hz can be divided into bands of equal contribution to speech intelligibility.

Ultimately, of course, the articulation-index concept has its foundation on an analysis of the sound pressures (signal and noise) that are produced at the listener's ear. In practice, it is expeditious to reduce all statements of sound pressure at the listener's ear to spectrum level (the sound pressure level in decibels is 0.0002 dyne per square centimeter for a 1-Hz bandwidth).

While much too complex to describe in detail here, speech intelligibility and quality are of major concern in the design of certain communications systems. The references listed contain considerable further information on this topic. See also **Telephony;** and **Voice and Sound Production.**

ARTIFICIAL HORIZON. A planar reflecting surface that can be adjusted to coincide with the astronomical horizon, i.e., can be made perpendicular to the zenith. This instrument is used, usually in conjunction with others, in observing celestial bodies. See also **Horizon (Astronomical); Horizon (Celestial);** and **Sextant.**

ARTIFICIAL HORIZON (Aircraft). Several ways of presenting to the pilot the information as to direction of true vertical have been used. The artificial horizon indicator shown in Fig. 1 is one acceptable means. Gyros furnish an excellent means for determining dynamic vertical. If a gyro is arranged with gimbals as shown in Fig. 2, and there is some method of detecting instantaneous or dynamic vertical by which

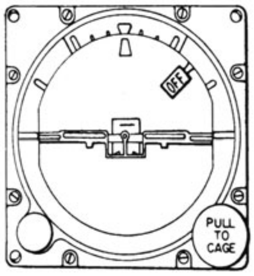


Fig. 1. Artificial horizon indicator.

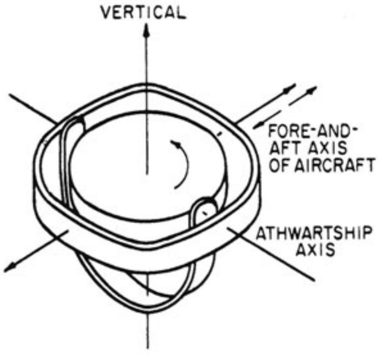


Fig. 2. Gyro in gimbals.

a torque may be applied to the gyro with its downward component being directed to the rising side of the gyro wheel (Fig. 3), the spin axis will move toward the dynamical vertical. As the rate of this movement toward vertical (called the erection rate) is a relatively slow action, of the order of a few degrees in a minute, the spin axis indicates a long-time average of dynamic vertical which usually is very close to true vertical.

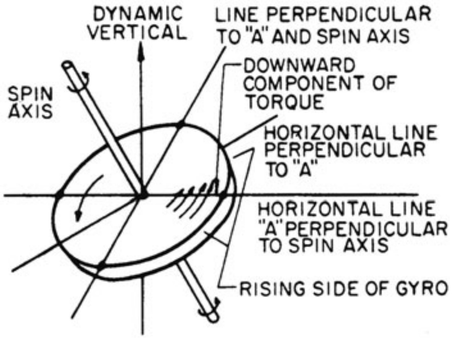


Fig. 3. Gyro vertical.

Often, the detecting means may be pendulous switches, usually electrolytic or mercury, which energize torque motors causing the gyro to erect. This system has many advantages in that the erection rate is externally adjustable if desired, and erection may be disconnected by means of switches actuated by rate gyros so that only when the dynamic vertical is near true vertical is it averaged by the gyro mechanism and a more exact true vertical indication is obtained.

If the gyro spin axis is maintained horizontal instead of vertical, it will tend to keep its spin axis pointed in one direction. By putting a degree calibration on the gyro, it forms a stable heading reference. Methods of drive, erection, and signal takeoff are similar for both gyro verticals and horizontals.

ARTIFICIAL INSEMINATION. Introduction of semen into the vagina by means of surgical instruments and procedures. Although sparingly used in *Homo sapiens*, artificial insemination is widely practiced in the breeding of horses and livestock. See also **Embryo;** and **In-Vitro Fertilization.**

ARTIFICIAL INTELLIGENCE. The concept of *artificial intelligence (AI)* probably dates back a number of centuries, but only within the last half-century has AI been complemented with the trappings of science. The term *cybernetics* was coined by Norbert Weiner, an American mathematician, in 1947 when referring to the underlying principles involved in communication and the techniques whereby information is translated into constructive performance. This area of scientific interest resulted from the technical attention given to so-called "electronic brains" utilized in automatic control mechanisms, such as military weaponry and bombsights, that were developed during World War II. The word *cybernetics* is seldom seen in the literature today, but it is closely akin to what is currently considered to be the field of artificial intelligence. Cybernetics targeted the similarities between the human nervous system and brain and those of manufactured control systems. Weiner also extended the application of cybernetics, particularly the concept of feedback, to social and economic systems.

Interest in AI was rekindled in the mid-1950s by military and defense contractors, notably of aircraft, in the United States and the United Kingdom. This period coincided with the first real and practical uses of digital computers (i.e., when the digital computer was considered by many scientists as the ultimate answer to numerous problems), including that of providing a machine or process with an ability to think, giving rise to the term *artificial intelligence.* During that initial period, comparatively little scientific thought was given to the fundamental *inadequacies* of the digital computer to mimick the human thought process. In retrospect, it is surprising to note that many brilliant people in the AI field failed to recognize early on that the "on-off" and "if-then" logic of the digital computer dealt with precise bits of data and was incapable on its own to provide those actions that we ascribe to the human brain. See accompanying table. Great pride was taken in playing and winning games of chess and backgammon by computer, whereas, in reality, these were poor demonstrations of the complete human intellect. It was not until the 1970s that AI specialists took serious note that AI was much more than a high-speed digital computer. Soon thereafter, much greater emphasis was put on gaining a better understanding of how the human brain works and on different forms of computers, including neural networks. As a shortcut measure, "copycat" measures, including expert systems and model-based systems, were taken and

SOME INTELLIGENCE OPERATIONS OF THE HUMAN BRAIN

Function	Action
Cogitation	The process of "mulling over" various possible choices to take in selecting the one best (mentally optimized) action or possibly a series of "best" actions, as compared with numerous, less-effective solutions. Cogitation improves with experience. This process calls on judgment and discernment.
Reasoning	The justification of a decision prior to final execution of a decision.
Deliberation	The unhurried weighing of information (of varying degrees of input accuracy) in terms of probable success or failure.
Speculation	The weighing of what might be termed chancy (auspicious, risky) pathways in an effort to maximize possible gain versus risk.
Cognition	A past-knowledge-based process of what may be good or inferior practice. A prime example of the thought process, advantage of which is taken in expert and model systems.
Approximation	Also based upon past knowledge through experience, approximation is the ability to estimate values and the probable result of using these values.
Weighting	Sorting out factors of serious importance versus other factors that may be at the noise level.
Interpolation	The creation of values for important, but missing, input data.

achieved some successes. Many such systems are in place as of 1994. These are discussed later in this article.

It was not until 1977, some 25 years after the inception of the AI concept, that some scientists vocalized their concern that the digital computer was not a complete answer. For example, I. Goldstein and S. Papert observed, "The fundamental problem of understanding intelligence is not the identification of a few powerful techniques, but rather the question of how to represent large amounts of knowledge in a fashion that permits their effective use and interaction . . .. The current point of view (1977) is that the problem solver (whether man or machine) must know explicitly how to use its knowledge—with general techniques supplemented by domain-specific pragmatic knowhow. Thus, we see AI as having shifted from a *power-based* strategy for achieving intelligence to a *knowledge-based* approach."

Recognition of these early shortcomings of AI research ultimately led to greater realism and the introduction of "copycat" systems. It also led to much more research on understanding the human brain and on how, in some way, the knowledge can be married with digital computers or to the development of *new concepts* in computing (possibly more aptly termed *decision-making devices*).

Since the late 1980s, some AI scientists have achieved a partial degree of success through the development of *fuzzy logic* computer software.

Fuzzy Logic

The human brain, unlike the digital computer, is equipped to store and and process vast multitudes of "shades of meaning." One of the developers of fuzzy logic, Pat Murphy (Omron Electronics, Inc.), observed recently, "By approximating the kind of unconscious reasoning people use to solve problems, fuzzy logic programming can help speed system design and development, simplify debugging and system modification, and make it possible to develop control programs for systems that, up to now, could not be effectively automated. The chief appeal of fuzzy logic is its ability to provide smoother, more human-like control in everything from auto-focus video cameras and microwave ovens to automating automobile transmissions and automatic braking systems."

Fuzzy logic software is becoming available for certain situations where it is necessary to deal with "ranges of values" or "degrees of normalcy," rather than with precisely specific values which, traditionally is the bailiwick of the digital computer, that is, the *crisp set*.

For example, the computer, unless prorgrammed with fuzzy logic, cannot distinguish a dispersion of measured values that may range between the "ideal, normal, or desired" and other values, which, for all practical reasons, may be "acceptably high" or "acceptably low," thus constituting a *range* of acceptable values. Common sense reasoning is not present and thus a range of acceptable values is rejected (not recognized).

This is explained further by Murphy, "In fuzzy set theory, an element's membership in a set is a matter of degree. An element may be a member in more than one set. Although this concept seems complex, it is based on a mode of categorization that people, computer unassisted, use on a regular basis." In some cases, this can easily spell the difference between profit and loss.

Fuzzy sets, membership functions, and production rules are the three primary elements of a fuzzy logic (or fuzzy inference) system. Fuzzy set theory originally was developed by Lotfi Zadeh (University of California, Berkeley) in 1965. In his efforts, Zadeh was attempting to reconcile control theory with the modeling of human experience and intuition. Zadeh reasoned that the rightness of conventional set theory made it impossible to account properly for the elements of vagueness, imprecision, and shades of gray that are commonplace in the real world.

In conventional set theory, upon which Aristotelian logic is based, elements are either a member of a given set or they are not. There is no middle ground.

The following points summarize the differences between conventional (traditional) computer programming and what may be termed unconventional (AI) programming:

1. Unconventional (AI) Computer Programming
   Symbolic processing—a program that works out relationships among objects.
   Heuristic search—exploratory, recognition of promising approaches to a problem; sometimes called the art of good guessing.
   Tolerates some incorrect answers
   Less than best answer usually acceptable
   Easy to enlarge, modify, and update
   Control structure usually separated from domain knowledge
2. Conventional (Traditional) Computer Programming
   Numeric processing
   Algorithms used—solution steps are explicit
   Correct answer required
   Difficult to modify
   Information and control integrated

Expert Systems

In an expert system, the experience of one or more specialists in a given field is worked into a computer program. The concept is to create a program that will furnish answers of the quality that represents the best suggestions that would emanate from seasoned professionals in the field, just as if they were "inside" the computer, so to speak. This requires extensive and a deeply probing effort on the part of the programmer or "knowledge engineer."

Over the years, a number of expert systems have been developed. These include applications in the fields of chemistry, geology, genetic engineering, medicine, military situations analysis, oil well drilling, among others. The simplest and most successful applications are classification programs, as exemplified by differential diagnosis in the medical field. Several of these programs are described in some detail in the references listed.

A specific example of an expert system used for self-tuning automatic controllers is given toward the end of this article.

Language Understanding

Much of the AI research pertaining to natural language understanding has targeted on the written word, rather than to attempt the deciphering of voice sound waves. The written word, of course, is very limiting from the start in such programs, simply because individuals put a lot of meaning into words by the way they utter them. In the very early days of AI, the suggestion was made to use the techniques of message decoding. Many attempts were made, including the first Russian/English translation program prepared by A. G. Oettinger. Even though commenced in the mid-1950s, little improvement had been made by 1966. During that year, the National Research Council's Automatic Language Processing Advisory Committee, which at one time had been highly enthusiastic, recommended that most of the funding for machine translation research be terminated. The conclusion—whatever it is that language encodes, it is not just a matter of words and definitions and vocabulary. Somewhere behind the surface structure of human language there is an enormous body of shared knowledge about the world, an acute sensitivity to the nuance and context, an intuitive insight into human goals and benefits. Any machine for translating between languages must first understand what is being said, which means, in essence, that the machine must know a great deal about the world beforehand. As pointed out by Yehoshua Bar-Hillel in 1960, a translation machine should not only be supplied with a dictionary, but with a universal encyclopedia as well.

A major coordinated effort at speech understanding was later sponsored by the Defense Advanced Research Projects Agency (DARPA) between 1971 and 1976. The results may be described as modest. One test, in addition to the military, of the success of computerized language translation, of course, is the marketplace. There obviously are many ready-made needs awaiting an effective system.

Robots and Image Analysis

Self-manipulation and guidance of robots and the image analysis problems (pattern recognition, etc.) that are present in machine vision, for example, make for natural opportunities in the application of AI technology. Some measurable process has been made. Check alphabetical index and, in particular, see articles on **Machine Vision;** and **Robotics.** Expert systems have been developed to self-tune automatic controllers.

Self-Tuning. Consider a heat exchanger that uses saturated steam to heat water that flows through its tube bundle. A simple control scheme senses the outlet water temperature and attempts to position the steam valve so that the actual water temperature equals the desired water temperature. Effects of both nonlinearities in the steam valve and the changing steam pressure can be reduced by using a second control loop to control the steam flow (known as self-adaptive control). Unfortunately, a fixed-parameter temperature controller has difficulty because of the nonlinear, time-varying behavior of the process. A change in the water flow rate changes the effective delay time and heat transfer characteristics of the process. Gradual fouling of the heat exchanger tubes also changes the performance dynamics over time. Good control performance at one operating condition can give way to very poor performance (overdamped or unstable response) at another operating condition. The need to monitor and change (tune) a controller that operates on a 24-hour basis, often 365 days in a year, is obviously a task that is a candidate for automating, but this is not possible to accomplish with a garden-variety controller. An expert system (or equivalent) is required.

In general, time-varying process dynamics, variable operating conditions, nonlinear process dynamics, and lack of expertise during control loop commissioning have all led to an interest in *self-tuning* controllers. Economic incentives are ultimately behind the thrust, since control loop performance directly affects product quality, energy consumption, product yield, production rate, pollution control, and plant safety.

The meaning of the term "self-tuning" is somewhat nebulous, especially when it is combined with the term "adaptive." Self-tuning or self-adaptive control imply the ability to learn about the closed loop process in real time. The controller initially tunes itself and remains tuned as the process dynamics and operating conditions change. The implied assumption underlying the theory is that the best present control strategy can be based upon past closed loop observations. Many schemes, such as dead time compensation, gain scheduling and feedforward, which have been labeled adaptive are actually preprogrammed adjustments, based upon measurable quantities. Such techniques are well suited to compensate for nonlinearities caused by actuator characteristics or easily modeled physical phenomena, and good design practice dictates that they should be used wherever possible. There is nothing to be gained by employing the complexity of self-adaptation to learn what is already known. There are many instances, however, where the causes of behavioral changes are either unknown, unobservable, or difficult to model. These are the situations to which self-tuning techniques are best applied.

Two entirely different self-tuning approaches have evolved and are commercially available, an expert system and process model approach. The expert system uses heuristics or rules of thumb. Its goal is to achieve a desired control loop response by incorporating tuning rules used by control engineers to manually tune controllers plus additional rules discovered during field tests. Tuning changes result directly from the process response without any need to mathematically model the process. The control loop response is expressed in terms that describe its pattern. Examples of such pattern features are peak heights, period, slopes, frequency content, integrated areas, zero crossings and other shape information. These are the same features that the eye detects when its scans a strip chart recording of the reference input (set point), measurement, and controller output. The discrete nature of the pattern characteristics easily combine with the tuning rules which are expressed in the IF-THEN-ELSE format. The control structure usually chosen is a discrete form of the Porportional + Integral + Derivative (PID) controller because of widespread industrial acceptance and a long history of tuning experience. See **Control Action.**

The model based self-tuning approach depends entirely upon a process model. Its goal is to achieve a desired control loop response by updating coefficients in a process model and using the coefficients to calculate the control parameters. If the model is appropriately constructed, the calculation is simple since the model coefficients can be directly used in the controller. The desired response is expressed as a transfer function which relates the measurement to the reference input and process load disturbance. The identified process model coefficients minimize the mismatch between the actual process response and the model response to measured inputs. This model based approach is flexible enough to accommodate a wide variety of parameter identification techniques and controller design strategies. The complexity of the chosen control structure often necessitates self-tuning since it is almost impossible to manually tune a controller having more than three adjustments.

Both self-tuning approaches have their own advantages and disadvantages. Although it is usually difficult to get universal agreement because of subjective reasons, the comparative issues are presented here to allow a look into the future. The advantages of the expert system arise since it is extremely robust and additional rules can be easily added. It is also easy to apply because it is designed to mimic manual tuning rules which are understood by control engineers and technicians. Further, it responds directly to the quality of the closed loop process response and does not require a process model. On the negative side, the expert system is engineered for a particular controller structure and its rule base makes it impossible to analyze mathematically. It prefers single event unmeasured loop disturbances and may be more dependent upon the nature of an arbitrary unmeasured disturbance.

The model based approach, on the other hand, has a rigorously defined performance criterion and mathematical analysis is possible. Also, the process model allows the flexibility of implementing different controller structures and may be used for process diagnostics such as detection of sensor or actuator failure. On the negative side, the model structure may not allow a match with the process and final actuator dead zone or backlash causes underestimation of process gain. An estimate of the process dead time is needed and the model sampling time is critical. Too short a sampling interval leads to a model with too many parameters and too long an interval results in sluggish control behavior. Rapid process changes and unmeasured load disturbances, even single event disturbances, may give problems. Finally, it is difficult to develop a weighting scheme for past errors and to factor the model into its controllable and uncontrollable (i.e., delay time and nominimum phase zeros) parts.

None of these problems are insurmountable, however, since many successful applications are installed and running. The controller of the future won't be a plug-in black box because new processes tend to continually require more sophisticated controls and, at the same time, specifications become tighter.

The controller of the future will, however, incorporate both the expert system technology and the model-based technology. The two approaches will become one as their unique characteristics are blended for the common goal of process control.

In an expert system, one must first formalize the rules used to manually tune controllers as part of the knowledge base. The art of manually tuning controllers based on pattern recognition has evolved over a number of decades. The control engineer either disturbs the closed loop by making a set point change or a load change, or else the engineer will wait for a natural disturbance. Based upon the closed loop response pattern to the upset, the controller tuning parameters are adjusted. This manual tuning procedure does not utilize a process model because, in general, processes are nonlinear, time varying, noisy, and they often have discontinuous regimes. As a result, processes are almost impossible to describe mathematically.

Figure 1 shows a block diagram of an expert system approach to self tuning. The knowledge-based rules are the formal implementation of the manual rules just described. The inputs for these rules are the set

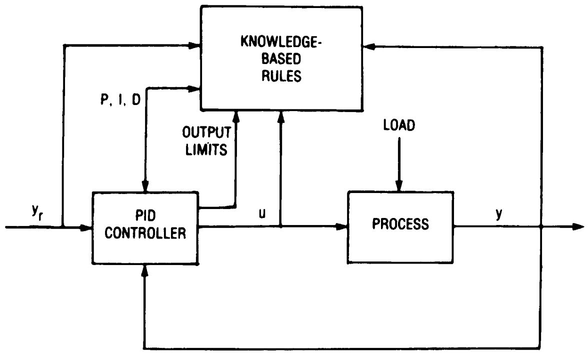


Fig. 1. An expert system approach to self-tuning a process controller. (*Foxboro.*)

point, measurement, controller output, controller output limits, and the current tuning parameters. Since a process model is not used, the tuning rules must be tailored to a particular controller structure, such as the common proportional + integral + derivative (PID) controller.

Pattern Recognition. Two hurdles must be overcome in building such an expert system. First, the adapter must recognize the pattern of the closed loop response. Second, after recognizing this pattern, it must be able to take proper control action. Figure 2 shows a transient response of error versus time for a load change. The dominant pattern features of this response are the peak heights shown as E1, E2, and E3. In a specific expert system configuration, these peak heights are normalized to define two variables, the overshoot and the damping. Note that the value of 1.0 for both these parameters indicates neutral stability and value of 0.0 represents an overdamped response. The damping is independent of the zero error line, whereas the overshoot depends strongly on this line. As shown in Fig. 3, another parameter is necessary to completely define the response pattern. A single process with different controller settings can produce two responses that are very similar in terms of overshoot and damping, but different in timing. The response period is the third pattern feature needed to distinguish the response. The period is the time between the first and third peaks and it is normalized with the integral term of the controller and with the derivative term of the controller to produce the ratios, integral/period and derivative/ period.

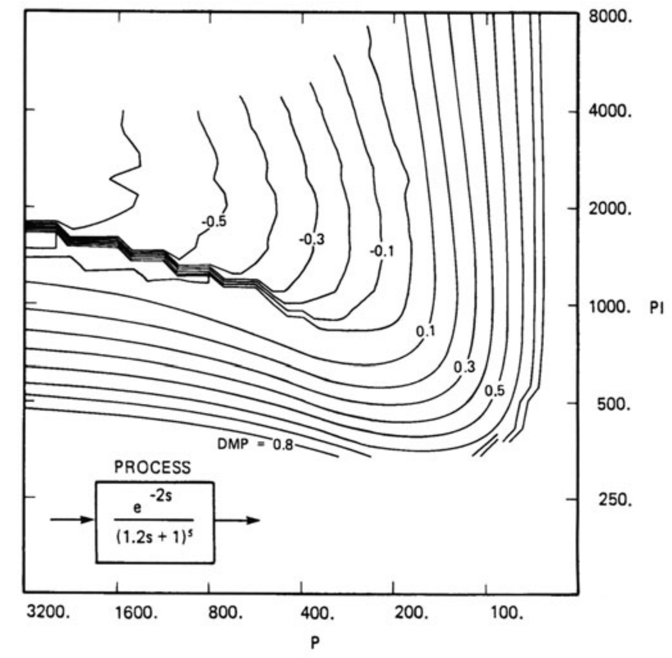


Fig. 4. Damping mapped into control space. (*Foxboro.*)

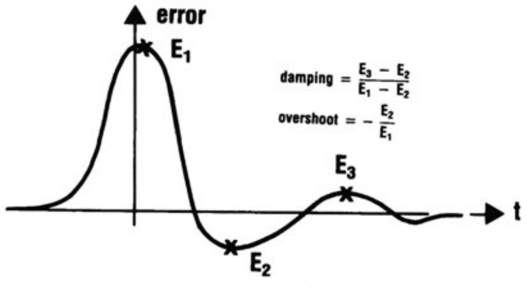


Fig. 2. Pattern recognition characteristics. (*Foxboro.*)

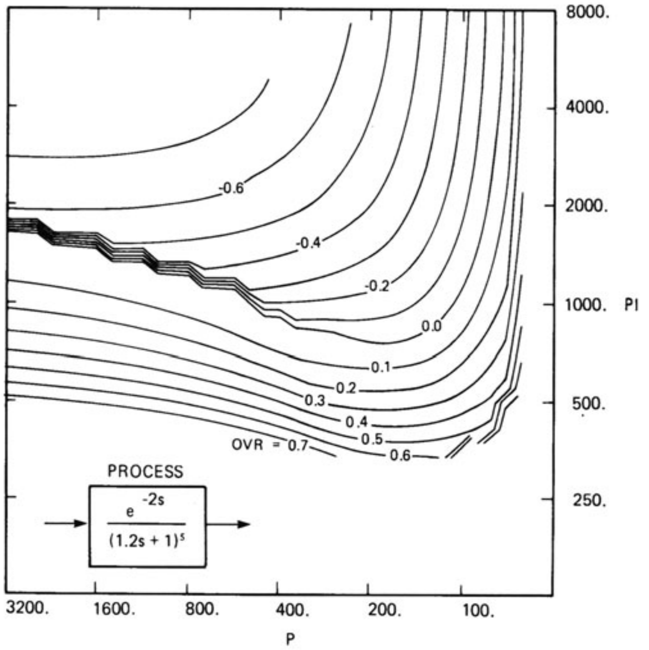


Fig. 5. Overshoot mapped into control space. (*Foxboro.*)

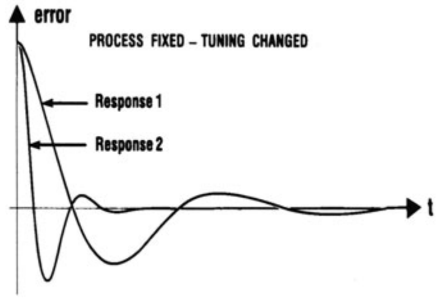


Fig. 3. Damping and overshoot are not independent. (*Foxboro.*)

These two time parameters are similar to the Ziegler and Nichols tuning ratios, but there are two important differences. First, this period includes the effect of the controller integral and derivative terms. This is significant because the integral and derivative terms of the controller change the closed loop period. The second difference is that the optimal integral and derivative ratios change depending upon the process characteristics. Processes having a large amount of dead time require smaller ratios whereas processes having more lag require larger ratios. The optimal integral and derivative ratios are adjusted automatically based upon the pattern of the process response.

Figures 4, 5, and 6 show contour plots of damping, overshoot, and integral/period for one specific process. The axes of these plots are the proportional term and the proportional term multiplied by the integral term. The contours represent lines of constant damping and overshoot. The similarity of the damping and overshoot contour shapes indicates that these two parameters are not independent and the same shape can be achieved from a number of different controller tuning settings. The knee in Figs. 4 and 5 represents the good tuning area. Note the damping and integral/period contour plots are orthogonal in this area, which suggests rapid convergence and unique controller tuning parameters produce the desired response pattern.

Knowledge-Based Rules. Up to this point, the response pattern is defined in terms of peak heights and period. These pattern features are normalized to produce damping, overshoot, integral/period, and derivative/period. This information is built into an expert system via knowledge-based rules as shown in the state diagram of Fig. 7. The quiet state corresponds to a mode where the measurement is close to the set point and new process information is not available. The adapter begins to watch the closed loop response when a process disturbance causes the

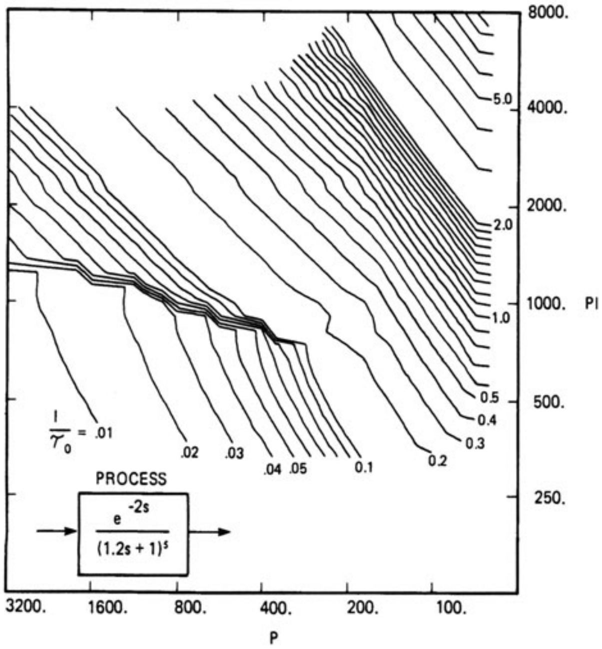


Fig. 6. Integral/period mapped into control space. (*Foxboro.*)

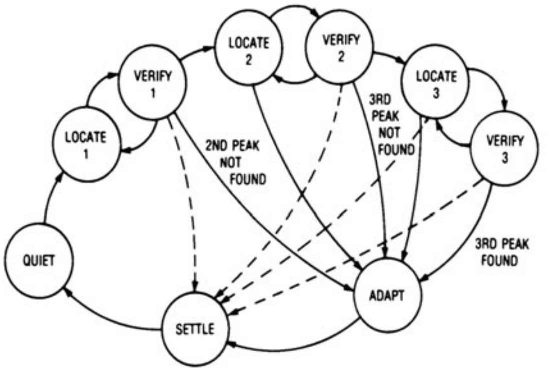


Fig. 7. State diagram. (*Foxboro.*)

measurement to drift away from the set point. Simultaneously, the controller, with its parameters fixed, takes action based upon the changing measurement. The adapter locates and verifies Peak 1, locates and verifies Peak 2, and locates and verifies Peak 3. The time it takes to observe these three peaks is highly process dependent. However, once the three peaks are observed, it quickly moves into the adapt state and adjusts the tuning parameters, if necessary.

The new tuning parameters are adjusted in two steps. First, the integral and derivative term are set based upon the desired integral/period and derivative/period ratios. Because of parameter interaction, the proportional term is adjusted to compensate for the integral and derivative changes. Second, the observed damping and overshoot are compared to the desired values. If distinct peaks have occurred and both damping and overshoot are less than the desired values, the proportional term is decreased. The amount of decrease depends upon either the difference between the desired and actual damping, or the difference between the desired and actual overshoot. Since damping and overshoot are not independent, the smallest (or most negative) difference is used. If distinct peaks are not detected, all tuning parameters are decreased by an amount that depends upon the desired damping or overshoot.

After the adapt state, the process is controlled with new tuning parameters. The adapter then goes through a transient state called settle and back to the quiet state where it is ready to repeat this procedure again as required. When the response is overdamped and peaks are not present, the rule base recognizes the lack of peaks and goes directly to the adapt state, as shown in Fig. 7. Also shown in Fig. 7, the dotted lines represent a large set point change that occurs during the response to a previous control loop disturbance. The set point change alters the response error pattern and causes the control algorithm to abort its current pattern search.

The vast majority of tuning rules lie within the adapt state. This state contains kernels of knowledge in the IF-THEN-ELSE format. The rules represent small blocks of knowledge that work together to produce a very robust adapter. There are no hidden assumptions about the process or the controller. Stability is assured by the direct performance feedback aspect of this design. If the loop is initially unstable, the adapter watches the oscillations and adjusts the controller's tuning parameters until the oscillation decays. The adapter moves the integral and derivative based on the period and primarily moves the proportional band based upon the peak heights and the desired response shape. Once proper tuning is achieved, subsequent process disturbances produce the desired response.

Additional Reading

Alexsander, I., Editor: "Neural Computing Architectures: The Design of Brain-Like Machines," MIT Press, Cambridge, Massachusetts, 1989.

Angel, L.: "How to Build a Conscious Machine," Westview Press, Westview, British Columbia, Canada, 1990.

Bartos, F. J.: "AI Now More Realistically at Work in Industry," *Control Eng.*, 90 (July 1989).

Brooks, R. A.: "New Approaches to Robotics," *Science*, 1227 (September 13, 1991).

Duhamel, M. C.: "Mind Over Machine," *Case Western Reserve University Magazine*, 11 (August 1990).

Eberts, R.: "Process Operator Task Analysis and Training," in *Process/Industrial Instruments & Controls Handbook* (D. M. Considine, Editor), 4th Edition, 8.27 McGraw-Hill, New York, 1993.

Gevarter, W. B.: "Introduction to Artificial Intelligence," *Chem. Eng. Progress*, 21 (September 1987).

Hansen, P. D.: "Techniques for Process Control" in *Process/Industrial Instruments & Controls Handbook* (D. M. Considine, Editor), 4th Edition, 2.21, McGraw-Hill, New York, 1993.

Herrod, R. A., and L. Tietz: "AI Enhances Control Tuning," *InTech*, 34 (March 1990).

Horgan, J.: "Word Games: Another Attempt to Create a Universal Parsing Machine," *Sci. Amer.*, 34 (October 1991).

Horning, B.: "Language Busters," *Technology Review (MIT)*, 50 (October 1991).

Kurzweil, R.: "The Age of Intelligent Machines," MIT Press, Cambridge, Massachusetts, 1990.

Margolis, H.: "Patterns, Thinking, and Cognition," University of Chicago Press, Chicago, Illinois, 1988.

Mercadal, D.: "Dictionary of Artificial Intelligence," Van Nostrand Reinhold, New York, 1990.

Murphy, P.: "Fuzzy Logic Smooths System Control," *Instruments & Control Systems*, 45 (March 1992).

Peterson, I.: "The Checkers Challenge," *Science News*, 40 (July 20, 1991).

Prerau, D.: "The Application and Benefits of Expert Systems," "The Application and Benefits of Expert Systems," *Access* (GTE Laboratories Incorporated), 5 (Spring 1992).

Russo, M. F., and R. L. Peskin: "Knowledge-Based Systems for the Engineer," *Chem. Eng. Progress*, 38 (September 1987).

SanViovanni, J. P., and H. C. Romans: "Expert Systems in Industry," *Chem. Eng. Progress*, 52 (September 1987).

Selfridge, O., and J. Vittal: "History/Overview—Artificial Intelligence," *Access* (GTE Laboratories Incorporated), 2 (Spring 1992).

Smith, M., and M. Abdel-Rahman: "The Impact of AI on Sensing Technology," *Sensors*, 16 (September 1991).

Staff: "Artificial Intelligence in the Office and on the Factory Floor," *Westinghouse Technology*, 2 (April 1989).

Staff: "A Conversation with AI Founder, Marvin Minsky," *Access*, (GTE Laboratories Incorporated), 10 (Spring 1992).

Stephanopoulos, G.: "The Future of Expert Systems in Chemical Engineering," *Chem. Eng. Progress*, 44 (September 1987).

Stone, R.: "The Education of Silicon Linguists," 854 (August 23, 1991).

Vittal, J., and O. Selfridge: "AI—The Next Ten Years," *Access* (GTE Laboratories Incorporated), 8 (Spring 1992).

Wallich, P.: "Trends in Artificial Intelligence: Silicon Babies," *Sci. Amer.*, 124 (December 1991).

Wallich, P.: "Rapid Recall," *Sci. Amer.*, 32 (July 1992).

Waterbury, R. C.: "Real-Time AI Meets Real World," *InTech*, 28 (August 1989).

ARTIFICIAL LINE. An artificial line is an electrical network consisting of resistance, inductance and capacitance so connected that it has the same characteristics (electrical) as the actual transmission line. Sometimes where the artificial line is not required to duplicate exactly the actual line the inductance or capacitance may be omitted. Such a line is very valuable for making laboratory tests as it makes possible connection at points corresponding to an actual line over a long distance. Artificial lines are also used in telephone and telegraph practice to balance actual lines to give desired operating characteristics in bridge type circuits.

ARTIODACTYLA (*Mammalia*). Hoofed animals which retain an even number of toes, the axis of the foot passing between the third and fourth digits. In older terminology, the term even-toed ungulates was used. Organization of the *Artiodactyla* is shown in the accompanying table, along with references to specific entries in this volume which describe the various families, subfamilies, and species found in the order of *Artiodactyla*.

For references, see **Mammalia**.

ASAFETIDA. The gum-resin exudate from the roots of a commonly occurring plant in the steppes region of Asia. Upon steam distillation, it yields a pale-yellow to orange-yellow liquid having a garliclike odor and slightly bitter, pungent taste. The resin from this plant (*Ferula assafoetida* L.) of the family *Umbelliferae* contains methylpropenyldisulfides, along with small amounts of vanillin. Asafetida has been used as

ARTIODACTYLA
(Even-toed Hoofed Mammals)

- ANTELOPINES **See Antelope.**
  - Horse-Antelopes (*Hippotraginae*)
    - Sabre-horned Antelopes (*Hippotragus*)
    - Rapier-horned Antelopes (*Aegoryx* and *Oryx*)
    - Screw-horned Antelopes (*Addax*)
  - Deer-Antelopes (*Alcelaphinae*)
    - Hartebeests (*Alcelaphus*)
    - Damalisks (*Beatragus* and *Damaliscus*)
    - Gnus (*Connochaetes* and *Gorgon*)
  - Marsh-Antelopes (*Reduncinae*)
    - Waterbucks (*Kobus*)
    - Lechwes (*Onotragus*)
    - Kobs (*Adenota*)
    - Reedbucks (*Redunca*)
    - The Rhebok (*Pelea*)
  - Blackbuck (*Antilopinae*)
  - Pigmy Antelopes (*Neotraginae*)
    - Klipspringers (*Oreotragus*)
    - Oribis (*Ourebia* and *Raphicerus*)
    - Sunis (*Nesotragus*)
    - The Beira (*Dorcatragus*)
    - Dik-Diks (*Madoqua* and *Rhynchotragus*)
    - Royal Antelopes (*Neotragus*)
  - Gazelles (*Gazellinae*)
    - Impalla (*Aepyceros*)
    - The Gerenuk (*Litocranius*)
    - The Dibatag (*Ammodorcas*)
    - The Springbuck (*Antidorcas*)
    - The Addra (*Addra*)
    - True Gazelles (*Gazella*)
    - Goat-Gazelles (*Procapra*)

- ANTILOCAPRINES **See Pronghorn Antelope.**
  - Pronghorn Antelope

- BOVINES **See Bovines.**
  - True Oxen (*Bovinae*)
    - Cattle (*Bos*)
    - Buffalo (*Bubalus, Syncerus,* and *Anoa*)
    - Bison (*Bison*)
  - Deer-Oxen (*Boselaphinae*)
    - The Nilghai (*Boselaphus*)
    - The Chousingha (*Tetraceros*)
  - Twist-horned Oxen (*Strepsicerosinae*)
    - Elands (*Taurotragus*)
    - The Bongo (*Böocercus*)
    - Kudus (*Strepsiceros*)
    - Bushbucks (*Tragelaphus*)
  - Duikers (*Cephalophinae*)
    - Common Duikers (*Sylvicapra*)
    - Forest Duikers (*Cephalophus*)
    - Blue Duikers (*Philantomba*)

- CAMELINES **See Camels and Llamas.**
  - Camels (*Camelus*)
  - Llamas (*Lama*)
  - The Vicuña (*Vicugna*)

- CAPRINES **See Goats and Sheep.**
  - Gazelle-Goats (*Saiginae*)
    - The Chiru (*Panthalops*)
    - The Saiga (*Saiga*)
  - Rock-Goats (*Rupicaprinae*)
    - The Goral (*Naemorhedus*)
    - Serows (*Capricornis*)
    - Chamois (*Rupicapra*)
    - Rocky Mountain Goat (*Oreamnos*)
  - Ox-Goats (*Ovibovinae*)
    - Takins (*Budorcas*)
    - The Muskox (*Ovibos*)
  - True Goats (*Caprinae*)
    - Tahrs (*Hemitragus*)
    - Markhors (*Capra falconeri*)
    - The Tur (*Capra caucasica*)
    - Ibexes (*Capra ibex, ...*)
  - Sheep (*Ovinae*)
    - The Aoudad (*Ammotragus*)
    - The Bharal (*Pseudovis*)
    - True Sheep (*Ovis*)
- CERVINES **See Deer.**
  - Musk-Deer (*Moschinae*)
  - Muntjacs (*Muntiacinae*)
  - True Deer (*Cervinae*)
    - Père David's Deer (*Elaphurus*)
    - Fallow Deer (*Dama*)
    - Axis Deer (*Axis*)
    - Red Deer (*Cervus*)
  - Hollow-toothed Deer (*Odocoileinae*)
    - White-tailed Deer (*Odocoileus*)
    - Marsh Deer (*Blastocerus*)
    - The Pampas Deer (*Ozotoceros*)
    - Guemals (*Hippocamelus*)
    - Brockets (*Mazama*)
    - Pudus (*Pudua*)
  - Moose (*Alcinae*)
  - Reindeer (*Rangiferinae*)
    - Eurasian Reindeer (*Rangifer tarandus*)
    - Caribous (*Rangifer arcticus, ...*)
  - Water-Deer (*Hydropotinae*)
  - Roe Deer (*Capreolinae*)
- GIRAFFINES **See Giraffe and Okapi.**
  - Giraffes (*Giraffinae*)
  - Okapis (*Palaeotraginae*)
- HIPPOPOTAMINES **See Hippopotamus.**
  - Common Hippopotamus (*Hippopotamus*)
  - Pigmy Hippopotamus (*Choeropsis*)
- SUINES **See Suines.**
  - Pigs (*Suidae*)
    - Eurasian Pigs (*Sus*)
    - African Bush-Pigs (*Potamochoerus*)
    - The Forest-Hog (*Hylochoerus*)
    - Wart-Hogs (*Phacochoerus*)
    - The Babirusa (*Babirusa*)
  - Peccaries (*Tayassuidae*)
- TRAGULINES **See Tragulines.**
  - Oriental Chevrotains (*Tragulus*)
  - Water-Chevrotains (*Hyemoschus*)

a flavoring in a number of food products, including nonalcoholic beverages, ice creams, candies, baked goods, and condiments, among others. Most countries consider the available fluid extracts and tinctures as GRAS (generally regarded as safe).

ASBESTOS. A group of impure magnesium silicate minerals which occur in fibrous form. Colors may be white, gray, green, or brown; sp gr 2.5; noncombustible.

Serpentine asbestos is the mineral chrysotile, a magnesium silicate. The fibers are strong and flexible. Spinning is possible with the longer fibers.

Amphibole asbestos includes various silicates of magnesium, iron, calcium, and sodium. The fibers are generally brittle and cannot be spun, but are more resistant to chemicals and to heat than serpentine asbestos.

Because asbestos has long been indicated in asbestosis (similar to silicosis) and, in more recent years, considered a carcinogen, several countries have issued regulations that restrict its use. For many years, asbestos was considered a desirable material for use in fireproofing, brake lining, gaskets, roofing, insulation, paint fillers, reinforcing agent in rubber and plastics, and in electrolytic diaphragm cells. Before using or considering asbestos for a product, local governmental regulations should be checked.

Additional Reading

Holden, C.: "Asbestos Regulations to be Re-Examined," *Science*, 1639 (March 27, 1992).

Mossman, B. T., et al.: "Asbestos: Scientific Developments and Implications for Public Policy," *Science*, 294 (January 19, 1990).

Rom, W. N., and A. Upton: "Asbestos-Related Diseases," *N. Eng. J. Med.*, 129 (January 11, 1991).

Stone, R.: "No Meeting of the Minds on Asbestos," *Science*, 928 (November 15, 1991).

Stone, R.: "Fiber Flap: Refractory Ceramic Fibers," *Science*, 1356 (March 13, 1992).

ASCARIS. Parasitic roundworms of relatively large size found in the intestines of humans and other animals.

ASCENT OF SAP. All of the organs of any terrestrial plant are dependent for their existence upon water absorbed from the soil. This water, which always contains traces of solutes and hence is often referred to as sap, moves in a generally upward direction through the plant. In some of the tallest known specimens of redwood trees (*Sequoia sempervirens*) the sap must ascend to heights exceeding 350 feet (105 meters) if the topmost branch is to be kept supplied with water. The upward movement of sap in plants occurs in the xylem, which in trees and shrubs corresponds to the wood. In the trunks or larger branches of trees sap movement is confined to a few of the outermost annual rings of wood. Sap movement occurs only through the vessels and tracheids of the woody tissue.

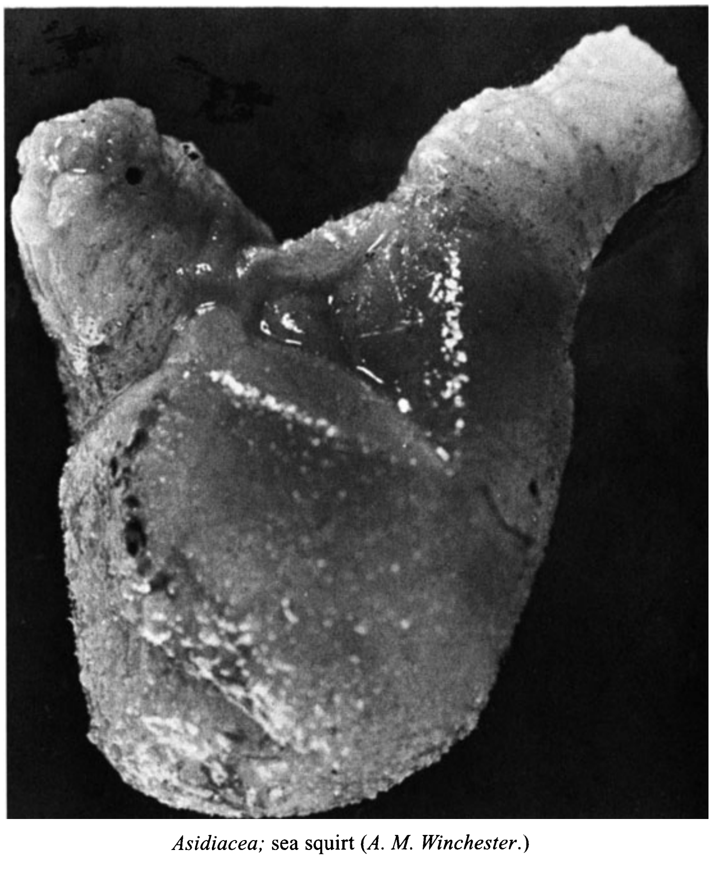
The earlier theories of the upward movement of sap in plants mostly invoked some vaguely conceived vital activity of the cells as furnishing the motive power for sap movement. Although the vessels and tracheids through which the water moves are dead, they are always in intimate contact with living wood parenchyma and wood ray cells and it is not inconceivable that these cells might in some way motivate the upward movement of sap. However, such theories receive very little support at the present time.

It is a common observation that sap may flow from the severed stems of many kinds of plants and that this flow ("bleeding") may continue for some time. This exudation of sap results from a pressure originating in the root called *root pressure*, and the exuded sap comes from the xylem tissues. Root pressures are also present in intact plants. For several reasons, however, root pressure can be considered only a secondary mechanism of water transport in plants. In the first place, there are many species in which the phenomenon does not occur. In the second place the magnitude of measured root pressures seldom exceeds two atmospheres which could not cause a rise of sap of more than about 60 feet (18 meters). In the third place known rates of sap flow under the influence of root pressure are inadequate to compensate for many known rates of transpiration. And finally, in woody plants at least, root pressures are usually present only in the early spring; during the summer period when transpiration rates and hence rates of sap movement are greatest, root pressures are negligible or nonexistent.

The principal mechanism motivating the ascent of sap in plants is thought by most present day botanists to be dependent upon the property of *cohesion* in water. The cohesive forces between water molecules are very great. The evaporation of water from the mesophyll cells of the leaf during transpiration results in the movement of water molecules into these cells from the xylem (water-conducting tissue) of the veins. The xylem of the leaves is continuous with that of the stems which in turn is continuous with that of the root system out almost to the very tip of every rootlet. The water is apparently present in the cells and vessels of the xylem as continuous threadlike columns. As water molecules pass out of these water columns into the mesophyll cells the threads of water become taut throughout the plant. Eventually a tension of considerable magnitude may be set up in them which is transmitted from the top to the bottom of the plant. The water columns can sustain this tension only because of the high cohesive force of water. When the water in the xylem of the younger roots passes into a state of tension, movement of water from the root cells into the xylem cells is induced. Loss of water from the root cells in turn causes absorption of water from the soil. Movement of water through the entire plant is thus brought about. Whenever transpiration rates are appreciable water does not, as a rule, enter the lower ends of the xylem ducts from adjacent root cells as fast as it passes from the upper ends into the mesophyll cells, hence the water is continuously under tension during periods of rapid transpiration and upward movement of sap. Calculations indicate that a cohesive force of between 30 and 50 atmospheres would be adequate to permit translocation of water to the very top of the tallest known trees by this mechanism. Experimentally determined values of the cohesive force of water are in excess of 300 atmospheres.

See also **Tree.**

ASCIDIACEA (*Chordata, Tunicata*). The tunicates, sea squirts, or ascidians (Ascidiacea), constituting a class of the subphylum *Tunicata*. They begin life as larvae which resemble tadpoles in form and later



Asidiacea; sea squirt (*A. M. Winchester.*)

become sessile animals invested in a covering called the test or tunic. Some species are solitary and others form colonies.

The class contains two orders:

Order *Enterogena*.

Order *Pleurogena*.

ASCITES. An abnormal accumulation of fluid in the abdominal cavity. Chronic heart failure will lead to blood being dammed up in the liver with resultant increase in its size. As a result of this congestion of the liver, the abdominal cavity may become filled with fluid (*ascites*). This fluid at times may enormously distend the abdomen. The fluid may be drained by a needle passed through the abdominal wall, or the kidneys may be forced to eliminate the fluid by the administration of diuretics.

Ascites is also common in patients with tuberculous infection of the peritoneum and in those with cancer which has spread through the peritoneal cavity. Tubercular ascites will disappear with eradication of the disease. Ascites is also a complication of portal cirrhosis; and occurs in connection with right-sided congestive heart failure.

ASCOMYCETES (Sac Fungi; Fungi). Many of the 40,000 species of fungi comprising the Ascomycetes are very common plants, but few are conspicuous. The great part of the species are small, often minute, while a few attain heights of 3 or 4 inches (7.6 or 10.2 centimeters), with a diameter of 1–2 inches (2.5 to 5.1 centimeters). Occasional individuals are even larger. All are characterized by the ascus, or spore-sac, commonly an elongate cylindrical body containing eight spores. In some species the ascus is spherical, or short cylindrical, while the number of spores may vary from two to many. Usually the asci are grouped together in a dense layer, called the hymenium. This may be composed entirely of asci, or may contain in addition numerous slender sterile filaments, called paraphyses. In some cases at least it seems the function of the paraphyses is to protect the asci, since the outer tip of each paraphysis is a flattened cap which partially covers the ascus. Ascomycetes are found wherever suitable food-yielding materials exist. Many species are parasites, living on living plants; among these are species of great economic importance. Other species are saprophytes, wood-destroying species being particularly numerous. See also **Fungus.**

The life-history of an Ascomycete comprises the mycelium composed of slender branching septate hyphae which penetrate throughout the substratum, and the fruiting stage in which the asci are formed. Two types of reproduction occur. One of these is the asexual type, in which asexual cells called conidia are cut off in various ways from the tips of hyphae, known as conidiophores. These conidia are single-celled spores which are disseminated by air currents. The other method of reproduction is sexual, and leads to the formation of asci. In *Pyronema confluens* this process has been carefully studied, and may be considered as typical in the main details for the process as it occurs in all the fungi of this class. The first step in this process is the formation of a multinucleate much-branched structure, which presently becomes septate. Some of the tips of this structure enlarge and become oögonia, called in this case ascogonia, while other tips become antheridia. From the oögonium a slender curved body called the trichogyne grows out. This is separated from the oögonium by a cross-wall. Since the oögonia and antheridia develop close together, the trichogyne comes in contact with the antheridium. All three bodies, oögonium, antheridium and trichogyne, are multinucleate. When the trichogyne comes in contact with the antheridium the walls between them at once break down, as does the wall between the trichogyne and the oögonium. The nuclei of the antheridium pass into the trichogyne, through it and into the oögonium. After this a new wall forms separating the trichogyne from the oögonium. In the oögonium the nuclei from the antheridium pair up with the nuclei of the oögonium, the nuclei of the trichogyne disintegrating early in the period of nuclear migration. Following the pairing of the nuclei in the ascogonium, coarse hyphae grow out from the latter. Into these the paired nuclei migrate. These coarse hyphae are the ascogenous hyphae, from which the asci eventually develop. In many Ascomycetes this process is considerably shortened, the ascogenous hyphae arising directly from the mycelium, no sex cells being formed; while other species have sex cells but no fusion, the oögonium alone developing.

The life cycles of the various ascomycetes are remarkably uniform, suggesting that they are all derived from a common ancestor. Two different views are held by botanists as to what the ancestral form may have been. According to one group, they are derived from red algae; favoring this view is the very great similarity in the development of the ascogonium and that of the carpospore formation in the algae; another favorable point is the presence of the trichogyne and the behavior of the antheridial nuclei. On the other hand, the other group holds that the ancestors of the Ascomycetes are to be found in the Phycomycetes, basing this contention on the similarity of the Phycomycete sporangium and the ascus, the latter being merely a sporangium in which the number of spores has been greatly reduced, becoming stable at eight in most species.

Many members of this class of fungi are of great importance to humans because of their destructive parasitic habit. A few species are of value as food, or in the production of foodstuffs, and other products used by humans.

Importance of Ascomycetes

Among the injurious species may be mentioned the Chestnut Blight fungus, *Endothia parasitica*, a disease probably introduced from China at the beginning of the twentieth century. In China the native chestnut trees had developed immunity; this the American trees did not have, so the fungus, which attacks the cambial tissue, was particularly destructive, nearly wiping out the native chestnut trees in a few years. See also **Chestnut Trees.** Another disease caused by an Ascomycete is the Brown Rot of stone fruits, caused by *Sclerotinia cinerea*. This fungus is particularly destructive in wet seasons. Often infected fruits become shriveled up and dry, in which condition they are known as "mummies." A large group of Ascomycetes are known as Powdery Mildews, because of the abundant conidiophores which are formed by the mycelium on the surface of the leaves of infected plants. Often these are so abundant as seriously to impair the functional efficiency of the leaf.

Another group of Ascomycetes contains species which are destructive and also those which are commercially of great value; these are the ubiquitous blue and green molds, species of *Aspergillus* and *Penicillium*. The destructive species attack foodstuffs everywhere, causing rotting and spoilage. Citrus fruits become covered with the bluish-green conidial masses; as does moist bread, pie crusts and many other foodstuffs. Species of the genus *Penicillium* give to Camembert and Roquefort cheese their characteristic properties. Other species of this group are the causal organisms for skin diseases of animals, including man. Another species of this genus, *Penicillium notatum*, is the source of the important drug penicillin. See also **Antibiotic.**

Among the largest of the Ascomycetes are species of truffles and morels, which are considered by mushroom fanciers to be particularly finely flavored. Truffles are fruit-bodies of the order *Tuberales*, and grow entirely underground. This makes it a matter of some difficulty to find them. Since they do not lend themselves to artificial cultivation, truffles must be sought in their wild habitat. To aid in locating them, dogs and pigs have been trained to find them using their superior sense of smell.

Another important Ascomycete is the genus *Claviceps*, which is parasitic on many grasses, including several cereal grains. This fungus forms a hard black sclerotium which is known as ergot, and which completely replaces the grain in the infected flower. The sclerotia are poisonous to livestock, causing the animals which have eaten them to become emaciated and covered with sores; another result is abortion in females. See also **Ergot.**

Another very important group of Ascomycetes is the Yeasts. See also **Yeasts and Molds.**

ASCORBIC ACID (Vitamin C). Infrequently referred to as the antiscorbutic vitamin and earlier called cevitamic acid or hexuronic acid, the present terms, *ascorbic acid* and *vitamin C*, are synonymous. Ascorbic acid was one of the first, if not the first nutrient to be associated with a major disease. Lind first described *scurvy* in 1757. However, this vitamin C deficiency disease had been recognized by Hippocrates in the

13th century and was a curse during the time of the Crusaders. In time of war, the disease killed untold numbers in armies, navies, and besieged towns. During the early days of the sailing ships, often requiring months between port calls accompanied by a lack of fresh food for long periods, the disease affected the crew as a plague. Scurvy was of some importance as recently as World War II. Currently, the disease is of prime concern in pediatrics. It is rarely seen in breast-fed children, but pasteurization of cow's milk degrades the vitamin and an addition to the diet of ascorbic acid must be provided for infants under 1 year of age.

Experimental scurvy was first produced by Holst and Frolich in 1907. About 6 months were required to produce scurvy experimentally, as individual susceptibility and the quantity of vitamin C previously stored in the body affects the onset of scurvy. The earliest sign of scurvy is usually a sallow or muddy complexion, a feeling of listlessness, general weakness, and mental depression. Soon the bones are affected and increasing pain and tenderness develop. Teeth easily decay and become loose and often fall out, while the gums bleed easily and are sore. Changes in the blood vessels occur, producing hemorrhages in different parts of the body. In infants, irritability, loss of appetite, fever, and anemia also occur. An infant between 6 and 12 months of age, who has not had sufficient intake of vitamin C (as from fruit juices, supplements, etc.) may show abnormal irritability and tenderness and pain in the legs, often accompanied by pain and swelling of joints (elbows and knees). Immediate administration of vitamin C is indicated in such cases.

In 1928, Zilva first described antiscorbutic agents in lemon juice, although the importance of fresh fruit or vegetables for preventing scurvy had been established a century or more earlier. Also in 1928, Szent-Györgyi isolated hexuronic acid (vitamin C) from lemon juice. In 1932, Waugh and King identified hexuronic acid as an antiscorbutic agent. Haworth, in 1933, established the configuration of hexuronic acid and, in that same year, Reichstein first synthesized hexuronic acid. Later in that year, Haworth and Szent-Györgyi changed the name of hexuronic acid to ascorbic acid.

In 1950, King et al, by the use of glucose labeled with radiocarbon in known positions, traced glucose through intermediate steps in the formation of ascorbic acid in plant and animal tissues, and then by using ascorbic acid with radiocarbon-labeled positions, it was possible to determine with considerable accuracy the metabolic distribution, storage, and chemical changes characteristic of the vitamin molecule. That experimentation made it clear that the carbon atoms in glucose or galactose all retain their original positions, along the carbon chain in the vitamin when it is formed biologically. No rupture or replacement in the chain during conversion was noted. It was also found that the synthesis can be considerably enhanced by feeding livestock small amounts of *chloretone* or any of a score or more of organic compounds. Reactions just described are indicated below.

Biological Role of Ascorbic Acid. Apparently all forms of life, both plant and animal, with the possible exception of simple forms, such as bacteria that have not been studied thoroughly, either synthesize the vitamin from other nutrients or require it as a nutrient. Dormant seeds contain no measurable quantity of the vitamin, but after a few hours of soaking in water, the vitamin is formed.

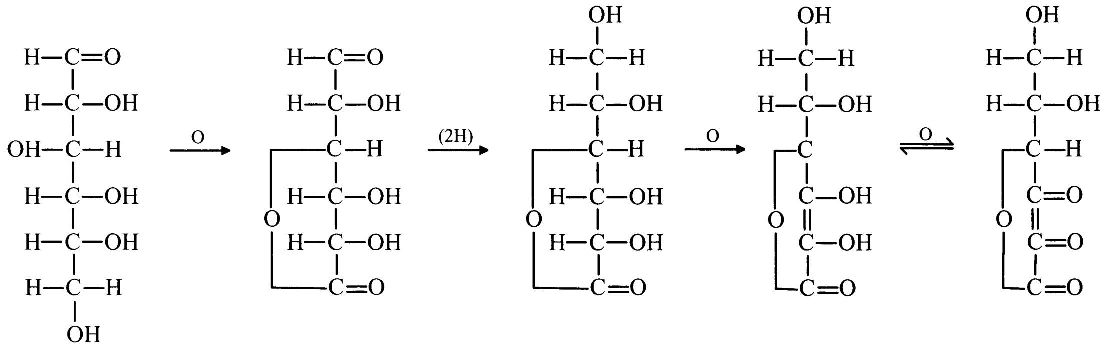
Ascorbic acid is easily oxidized to dehydroascorbic acid. The latter is less stable than ascorbic acid and tends to yield products, such as oxalate, threonic acid, and carbon dioxide. When administered to animals or consumed in foods, dehydroascorbic acid has nearly the same antiscorbutic activity as ascorbic acid, and it can be quantitatively reduced to ascorbic acid.

In its biochemical functions, ascorbic acid acts as a regulator in tissue respiration and tends to serve as an antioxidant in vitro by reducing oxidizing chemicals. The effectiveness of ascorbic acid as an antioxidant when added to various processed food products, such as meats, is described in entry on **Antioxidants.** In plant tissues, the related glutathione system of oxidation and reduction is fairly widely distributed and there is evidence that electron transfer reactions involving ascorbic acid are characteristic of animal systems. Peroxidase systems also may involve reactions with ascorbic acid. In plants, either of two copper-protein enzymes are commonly involved in the oxidation of ascorbic acid.

In animal tissues, it is easily demonstrated that, as the vitamin content of tissues is depleted, many enzyme systems in the body are decreased in activity. Full explanation of these decreased activities still requires further research. In the total animal and in isolated tissues from animals with scurvy, there is an accelerated rate of oxygen consumption even though the animal becomes very weak in mechanical strength and many physiologic functions are disorganized. With the onset of scurvy, the most conspicuous tissue change is the failure to maintain normal collagen. Sugar tolerance is decreased and lipid metabolism is altered. There is also marked structural disorganization in the odontoblast cells in the teeth and in bone-forming cells in skeletal structures. In parallel with the foregoing changes, there is a decrease in many hydroxylation reactions. The hydroxylation of organic compounds is one of the most characteristic features disturbed by a vitamin C deficiency. These reactions relate to the vitamin's regulation of respiration, hormone formations, and control of collagen structure.

A partial list of physiological functions that have been determined to be affected by vitamin C deficiencies include (1) absorption of iron; (2) cold tolerance, maintenance of adrenal cortex; (3) antioxidant; (4) metabolism of tryptophan, phenylalanine, and tyrosine; (5) body growth; (6) wound healing; (7) synthesis of polysaccharides and collagen; (8) formation of cartilage, dentine, bone, and teeth; and (9) maintenance of capillaries.

Requirements. Species known to require exogenous sources of ascorbic acid include the primates, guinea pig, Indian fruit bat, red vented bulbul, trypanosomes, and yeast. Species capable of endogenous sources include the remainder of vertebrates, invertebrates, plants, and some molds and bacteria. Estimates of requirements of vitamin C by humans has been approached in several ways: (1) direct observation in human studies; (2) analogy to experimentation with guinea pigs; (3) analogy to experimental studies in monkeys and other primates; and (4) analogy to animals, such as the albino rat, that normally synthesize the vitamin in accordance with physiological need. It is relatively easy to maintain intakes at recommended levels by use of mixed practical dietaries that include nominal quantities of fresh, canned, or frozen vegetables or fruits. Generally, ascorbic acid is considered as nontoxic to humans. Possible exceptions include kidney stones (in gouty individuals); inhibitory in excess doses on cellular level (mitosis inhibition) possible damage to beta-cells of pancreas and decreased insulin production by dehydroxyascorbic acid.



D-glucose D-glucuronic acid lactone L-gulonic acid lactone L-ascorbic acid L-dehydro ascorbic acid

Distribution and Sources. Natural sources of vitamin C include the following:

High ascorbic acid content (100–300 milligrams/100 grams)
  Broccoli, Brussels sprouts, collards, currant (black), guava, horseradish, kale, parsley, pepper (sweet), rose hips, turnip greens, walnut (green English)
Medium ascorbic acid content (50–100 milligrams/100 grams)
  Beet greens, cabbage, cauliflower, chives, kohlrabi, lemon, mustard, orange, papaya, spinach, strawberry, watercress
Low ascorbic acid content (25–50 milligrams/100 grams)
  Asparagus, bean (lima), cantaloupe, chard, cowpea, currant (red and white), dandelion greens, fennel, grapefruit, kumquat, lime, loganberry, mango, melon (honeydew), mint, okra, onion (spring), passion fruit, potato, radish, raspberry, rutabaga, soybean, spring greens, squash (summer), tangerine, tomato, turnip

For the species where the ascorbic acid is synthesized endogenously, the precursors include *d*-mannose, *d*-fructose, glycerol, sucrose, *d*-glucose, and *d*-galactose. Intermediates include uridine diphosphate glucose, *d*-glucuronic acid, gulon acid, *l*-gulonolactone, (Mn^{2+} cofactor). Production sites in animals are the kidney and liver in most instances. In rat, it is the intestinal bacterial supply. In plants, the production sites are found in green leaves and fruit skins. Cell sites include microsomes, mitochondria, and golgi.

Supplements. Commercially available ascorbic acid still includes isolation from natural sources, such as rose hips, but large-scale production will involve the microbiological approach, i.e., *Acetobacter suboxidans* oxidative fermentation of calcium *d*-gluconate; or the chemical approach, i.e., the oxidation of *l*-sorbose.

Bioavailability of Ascorbic Acid. The general causes of reduced availability of vitamin C include damage to adrenal cortex, presence of antagonists, and food preparation practices (oxidation, storage, leaching, cooking). Excepting the use of supplements, the almost universal requirement for fresh foods as a source of vitamin C is readily explained by the sensitivity of the vitamin to destruction by reaction with oxygen. This is accelerated by the presence of minute quantities of enzymes that occur in most living tissues, in which copper or iron is combined with a protein to form a catalyst for the oxidation reaction. Other chemicals, such as quinones or high-valence salts of manganese, chrominum, and iodine can also oxidize the vitamin readily in aqueous solutions. Most of these reactions increase rapidly in proportion to exposure to air and rising temperature. In the dry crystalline state, however, and in many dried plant tissues, particularly if acidic in reaction, the vitamin is quite stable at room temperature over a period of several months.

Freshly cut oranges or their juices may be exposed in an open glass for several hours without appreciable loss of the vitamin because of the protective effect of the acids present and the practical absence of enzymes that catalyze its destruction. In potatoes, when baked or boiled, there is a slight loss of the vitamin, but if they are whipped up with air while hot, as in the production of mashed potatoes, a large fraction of the initial vitamin content usually will be lost. In freezing foods, it is common practice to dip them in boiling water or to treat them briefly with steam to inactivate enzymes, after which they are frozen and stored at very low temperatures. In this state, the vitamin is reasonably stable. Vitamin C degradation in dehydrated food systems is described shortly.

Factors which increase the bioavailability of ascorbic acid include the presence of antioxidants and synergists in the diet.

Numerous studies have been conducted concerning vitamin C degradation during food processing, including dehydrated food systems. In the latter, the degradation is dependent upon water activity, moisture content, and storage temperature (for example, in containers with no headspace). Ascorbic acid destruction is dramatically increased in the presence of oxygen.

ASEXUAL REPRODUCTION. In asexual reproduction, a part of the parent organism becomes an organism identical with the parent.

Different types of asexual reproduction can be found in the different groups of plants. In bacteria, fission, a simple splitting into two equal parts, is the asexual method of reproduction. In green algae the characteristic method is by the formation of zoospores, which are unicellular motile bodies formed from the protoplast of a single cell. Each zoospore, after swimming around for a time, becomes quiet and grows to form a new individual of the parent type. In flowering plants there are many methods of asexual reproduction, some of great importance to humans. A common type is found in the strawberry, where long slender branches grow out from the short stem of the plant and take root at their tip. There a new plant forms. This reproductive structure is called a stolon. Many other plants produce similar branches which run along the surface of the ground (runners) or just beneath it (rhizomes), and send up one or more new plants from the nodes. Other plants, like the tiger lily, bear small buds in the axils of their leaves: the buds or bulblets are easily detached and readily grow to form new plants. Similar bulblets are formed at the base of many bulbs. Another very common method of asexual reproduction is through the formation of suckers, branches formed at the base of the parent plants, and gradually growing to replace them. Growers propagate date palms, pineapples, and bananas, for example, by means of suckers. The familiar potato tuber, the swollen tip of a rhizome, is another type of asexual reproduction which is of tremendous economic value.

All the methods so far enumerated are from the stem of the plant. However, any part of the plant may be a means of asexual renewal. Many plants are known which reproduce by means of the leaves. Several of these are now cultivated extensively as objects of beauty or of curiosity. For example, the African violet, *Saintpaulia*, and many of the ornamental begonias are readily propagated by leaves. Less widely known but equally interesting are species of *Bryophyllum* and *Kalanchoe*. In the notches of the leaves of these plants, while still attached to the parent plant or after being severed therefrom, tiny plants readily form and grow.

Many plants readily root when cut up into segments. Most people are familiar with the habit of the willow twig of striking root and growing when stuck in the ground. Equally well known is the geranium cutting, which is merely a branch removed from the parent plant and placed in a favorable environment.

Asexual reproduction is of much economic importance. New and improved forms of plants are constantly being made: by asexual means they are reproduced in the great quantities necessary for commercial use. Without such reproduction their formation in quantity would be practically impossible. (See **Grafting and Budding**.)

Asexual reproduction is also found in animals although it is restricted to some of the simpler phyla. Fission, a splitting in half, occurs in most of the *Protozoa*, such as *Amoeba* and *Paramecium*. In the class of *Protozoa* known as *Sporozoa*, one cell may break up into many reproductive bodies which may be called *merozoites* or *sporozoites*. The nucleus within a cell first divides many times and a plasma membrane forms around each nucleus and the surrounding cytoplasm, and the original cell breaks open liberating the "spores." The malarial parasite, *Plasmodium*, is a good example; merozoites are formed in red blood cells and sporozoites are formed in a mosquito.

Budding is a means of asexual reproduction found in sponges and in certain coelenterates, such as Hydra. Many of the simpler metazoans also have the power of regeneration; if cut into two or more parts, each part can grow back the missing parts, thus forming two or more complete animals. See Figs. 1 and 2.

Fig. 1. Budding hydra. (*A. M. Winchester.*)

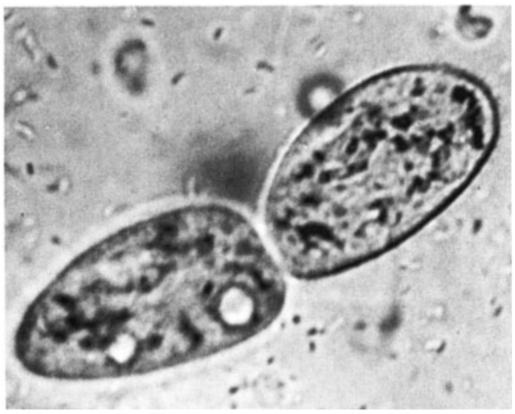


Fig. 2. Fission in paramecium. (*A. M. Winchester.*)

ASH TREES. The timber and shade trees called *ashes* are members of the family *Oleaceae* (olive family); whereas the mountain ashes are members of the family *Rosaceae* (rose family). The so-called prickly ash (toothache-tree), the southern prickly ash, and the hop tree (sometimes referred to as the wafer ash) are members of the family *Rutaceae* (citrus family).

The olive ashes are of the genus *Fraxinus* and are deciduous trees (a few may be considered shrubs). Important species include:

Afghan ash	*Fraxinus xanthoxyloides*
Biltmore ash	*F. biltmoreana*
Black ash (or hoop ash)	*F. nigra*
Blue ash	*F. quadrangulata*
Carolina ash (or water ash)	*F. caroliniana*
Chinese ash	*F. chinensis*
European common ash	*F. excelsior*
Flowering ash (or Marie's ash)	*F. mariesii*
Green ash	*F. pennsylvanica lanceolata*
Manna ash	*F. ornus*
Narrow-leaved ash	*F. angustifolia*
Oregon ash	*F. latifolia*
(Oxycarpa)	*F. oxycarpa*
Pumpkin ash	*F. profunda*
Red ash (or brown or river ash)	*F. pennsylvanica*
Texas ash	
(Tomentosa)	*F. tomentosa*
Velvet ash	*F. velutina*
Wafer ash (or hop tree)	*Ptelea trifoliata*
Weeping ash	*F. excelsior latifolia*
White ashF. texensis	*F. americana*

Blue ash located at Danville, Kentucky. (*Kentucky Division of Forestry.*)

The mountain ashes are of the genus *Sorbus* and are deciduous trees or shrubs. Important species include:

American mountain ash	*Sorbus americana*
European mountain ash (or Rowan)	*S. aucuparia*
Western mountain ash	*S. scopulina*
Notable Asian Species	
Embley mountain ash	*S. discolor*
Folgner mountain ash	*S. folgneri*
(Harrowiana)	*S. harrowiana*
(Hupehensis)	*S. hupehensis*
(Insignis)	*S. insignis*
Japanese mountain ash	*S. commixta*
Kashmir mountain ash	*S. cashmiriana*
(Pohuashaenensis)	*S. pohuashaenensis*
Sargent's mountain ash	*S. sargentiana*
Vilmorin's mountain ash	*S. vilmorinii*

As will be noted from the accompanying table of record ash trees, the white ash is a large tree and valuable for its timber. The white ash prefers moist and rich soil in a cool woods environment and particularly favors locations along rivers. The tree ranges from Newfoundland and Nova Scotia westward into Ontario and Minnesota, and southward to northern Florida and southwestward into Oklahoma and Texas. The tree frequently occurs scattered but there are large concentrations in Maine. The wood is excellent for furniture, interior finish, and implements of various types. The color ranges from pale to medium brown. The grain is close although not considered fine-grained. Commercial white ash in the green condition has a moisture content of about 43% and weighs 48 pounds per cubic foot (767 kilograms per cubic meter). After air-drying to 12% moisture content, the wood weighs 41 pounds per cubic foot (657 kilograms per cubic meter) and 1,000 board-feet (2.36 cubic meters) of nominal sizes weigh 3,420 pounds (1551 kilograms). Crushing strength of the green wood with compression applied parallel to the grain is 4,060 psi (28 MPa); of the dried wood, it is 7,280 psi (50.2 MPa). The tensile strength of the green wood with tension applied perpendicular to the grain is 580 psi (4 MPa); of the dry wood, 850 psi (5.8 MPa).

The leaves of the white ash are compound and from 7 to 12 inches (18 to 30 centimeters) long. They are of a lusterless light green, lighter and a silverish-green underneath. The flowers are green and without petals. The fruit occurs in clusters and is winged. The seed chamber is about 3/8-inch (0.9 centimeter) long and remains on bare branches well into mid-winter.

The green ash prefers the peripheries of streams and damp lowlands. It is found from Vermont southward along the mountains as far as northern Florida. Its range reaches westward to the eastern foothills of the Rocky Mountains. As shown by the accompanying table, the pumpkin ash attains great heights, normally between 60 and 100 feet (18–30 meters) for average trees. It is a slender tree with a trunk diameter usually of 3 to 4 feet (0.9 to 1.2 meters). The tree has a distinctively swollen appearing trunk at the base, giving rise to its unusual name. This species likes swamps and wet areas around ponds. It ranges from western New York west to Missouri and Arkansas and southward to Florida.

The black ash, also known as hoop ash, is also a slender, very tall tree, averaging from 50 to 65 feet (15 to 19.5 meters) in height and a trunk diameter of only 1 to 2 feet (0.3 to 0.6 meter). This tree also prefers swampy country. The blue ash is valuable for timber and is found generally in the western and southwestern states. See accompa-

RECORD ASH TREES IN THE UNITED STATES[1]

Specimen	Circumference[2] (inches)	Circumference[2] (centimeters)	Height (feet)	Height (meters)	Spread (feet)	Spread (meters)	Location
GENUS: *Fraxinus*							
Berlandier ash (1974) (*Fraxinus berlandierana*)	194	493	44	13.4	40	12.2	Texas
Black ash (1984) (*F. nigra*)	99	251	155	47.2	108	32.9	Michigan
Blue ash (1985) (*F. quadrangulata*)	199	505	60	18.3	74	22.6	Kentucky
Carolina ash (1988) (*F. caroliniana*)	56	142	48	14.6	42	12.8	Virginia
Green ash (1981) (*F. pennsylvanica*)	242	615	131	39.9	121	37.0	Michigan
Lowell ash (1974) (*F. anomala*)	47	119	50	15.2	16	4.9	Arizona
Oregon ash (1975) (*F. latifolia*)	263	668	59	18.0	45	13.7	Oregon
Pumpkin ash (1977) (*F. profunda*)	219	556	86	26.2	84	25.6	Virginia
Singleleaf ash (1973) (*Fraxinus anomala*)	19	48	24	7.3	23	7.0	Colorado
Texas ash (1984) (*Fraxinus texensis*)	72	183	42	12.8	35	10.7	Texas
Two-petal ash (*F. dipetala*)	35	89	34	10.4	28	8.5	California
Velvet ash (1971) (*Fraxinus velutina*)	140	356	66	20.1	65	19.8	Arizona
White ash (1983) (*Fraxinus americana*)	304	772	95	29.0	82	25.0	New York
GENUS: *Sorbus*							
American mountain ash (1979) (*Sorbus americana*)	80	203	62	18.9	40	12.2	West Virginia
European mountain ash (1984) (*S. aucuparia*)	73	185	39	11.9	38	11.6	Wisconsin
Showy mountain ash (1968) (*S. decora*)	57	145	58	17.4	32	9.6	Michigan
Sitka mountain ash (1981)	19	48	50	15.2	18	5.5	Oregon

[1] From the "National Register of Big Trees," The American Forestry Association (by permission).

[2] At 4.5 feet (1.4 meters).

nying photo. It can attain a height of close to 100 feet (30 meters), with a trunk diameter up to about 30 inches (76 centimeters), making it a slender tree. This tree is found in damp woods and prefers limestone hills. The tree occurs throughout middle America, ranging from Michigan southward into Alabama and Arkansas.

The American mountain ash is a very pretty tree, seldom exceeding 20–25 feet (6 to 7.5 meters) in height, with a trunk of from 8 to 15 inches (20.3 to 38 centimeters) in diameter. At high altitudes, it may be reduced to the status of a shrub. The tree flowers white in early spring. The fruit ranges in color from a coral red to deep scarlet. In nature, the tree is found in cool woods and along river banks, and also in swamps. The range is from Newfoundland west to middle-Canada and southward through the region of the Great Lakes and into Tennessee. The tree occurs on the mountain slopes in New England's White and Green Mountains and all along the Alleghanies as far south as North Carolina. The tree is used in landscaping.

The European mountain ash, sometimes called Rowan Tree, is used widely in parks and gardens in North America. The tree has narrow, oblong leaflets, the undersides covered with a white, hairy down. The fruit is of a bright scarlet color and is about 3/8-inch (0.9 centimeter) in diameter. There is considerably more variety in the European ashes with different colorations of bark and shape of leaves. There is a weeping ash (*Fraxinus excelsior* "Pendula") and the manna ash (*F. ornus*) which occurs in Southern Europe is known for its showy cream-colored flowers. The *F. excelsior* "Diversifolia" has a dandelion-color bark.

Several mountain ashes are native to Asia, including *Sorbus hupehensis* and *S. cashmiriana*. The *S. discolor* is well known as an excellent street tree because of its branches which point sharply upward. It has attractive large red leaves and fruit. The Chinese mountain ash, *S. vilmorinii*, is well known for its beauty of autumn coloration, with a full range from orange to purple.

The whitebeams are also of the genus *Sorbus* and thus closely related to the mountain ashes. A few highlights of the whitebeams include: *S. alinfolia* is found in Japan, crimson fruit, red-orange autumnal colors, attains a height of about 25 feet (7.5 meters); *S. aria*, a whitebeam of Europe, fragrant dull-white flowers, red fruit, brown autumnal colors, attains a height of from 30 to 45 feet (9 to 13.5 meters); *S. cuspidata*, a whitebeam of the Himalayas, white flowers, round, green fruit, large ovate leaves, attains a height up to 35 feet (10.5 meters); *S. domestica*, the service tree, occurs in southern Europe as well as western Asia and north Africa—pinnate leaves, small flowers, bears a pear-shaped fruit, may reach a height of 60 feet (18 meters); *S. hybrida*, also known as the bastard service tree, occurs in Scandinavia, white flowers, red fruit, reaches a height of 20 to 40 feet (6 to 12 meters); *S. intermedia*, the Swedish whitebeam, occurs in northwestern Europe, off-white flowers, clustered oval fruit of red coloration, attains height of about 20 feet (6 meters); *S. latifolia*, known as the service tree of Fontainebleau, ranges from Portugal to southern Germany. Rises to about 60 feet (18 meters), white flowers, brown fruit, shaggy bark; and *S. thibetica*, a whitebeam of Tibet.

For references see **Tree.**

ASM (International). The American Society for Materials is an organization with the mission of gathering, processing, and disseminating technical information pertaining to the understanding and application of engineered materials, their research, design, manufacture, and use, stressing economic and social benefits. These objectives are accomplished by way of a global information-sharing network of interaction among members in forums, meetings, and educational programs, through publications, such as the monthly *Advanced Materials and Processes,* and by use of electronic media. Headquarters of the ASM (International) is in Materials Park, Ohio.

ASPARAGUS. Of the family *Convallariaceae* (lily-of-the-valley family), genus *Asparagus*, the familiar asparagus plant is one of about 125 species in the genus. The plant is native to temperate and tropical regions of the Old World, but widely cultivated in suitable climes throughout the world. Members of the genus are characterized by having the leaves reduced to minute scales or bristles, while small, often very leaflike branches called cladophylls function as leaves. The flowers are small, yellowish or white in color, and the fruit is a berry, often brightly colored. *Asparagus officinalis*, a native of the marshes of Europe, is the cultivated garden form with thick and fleshy young stems. Other species are widely grown for their delicate beauty, as the familiar Asparagus ferns, *Asparagus plumosus* and *Asparagus sprengeri* (which are not properly ferns at all), and the florist's smilax, *Asparagus asparagoides*. See accompanying figure.

Bunches of asparagus placed on damp moss in flat to keep them fresh for market. (*USDA photo.*)

ASPARAGUS BEETLE (*Insecta, Coleoptera*). An introduced European beetle, *Crioceris asparagi*, which is sometimes an important pest on asparagus. A related species, also introduced from Europe, is known as the 12-spotted asparagus beetle. They are held in check by handpicking and by dusting plants with lime or other control chemicals.

ASPEN TREES. See **Poplar Trees.**

ASPERGILLOSIS. An infection produced by fungi of the genus *Aspergillus* of which *A. fumigatus* is the main infecting species, with *A. flavus* causing an invasive disease and *A. niger* an intracavity fungus ball of aspergilloma. The group of fungi is of low pathogenicity for humans unless resistance is overcome by an overwhelming inoculum, debilitating illness, or in immunocompromised individuals. There is no predisposition by race, age, or sex to *Aspergillus* infections.

Aspergilli are usually found as saprophytes on decaying vegetation, and pigeon excreta have also been found to harbor the organism. Marijuana also provides a site of growth and aspergillus precipitins have been found in the sera of more than 50% of marijuana smokers. The organisms assume a mycelial form both in culture and infected tissues.

Primary infections of the lung sometimes develops after inhalation of massive numbers of spores from mycelia growing on grain. Secondary pulmonary infection may be superimposed on tuberculous cavities, bronchiectases, and the like. The destructive paranasal granuloma normally caused by *A. flavus* is most often seen in the tropics or in patients who originate from those areas.

Allergic bronchopulmonary aspergillosis is associated with persistent endobronchial growth of the fungus—usually *A. fumigatus*. Symptoms of reversible airway obstruction are seen in early cases which develops into breathlessness and a chronic productive cough as the bronchial damage increases; mycelia of the fungus may be found in the sputum in such cases and x-rays show scattered linear shadows in the peripheral lung fields. Treatment is difficult, particularly in the late cases. Removal of the organism from the airways using antifungal drugs is normally of only temporary benefit. Therapy is therefore aimed at the inflammatory response by use of bronchodilators or corticosteroids.

The development of a fungal ball (aspergilloma) in a preexisting pulmonary cavity frequently presents no symptoms apart from intermittent cough associated with hemoptysis of varying severity. The condition is diagnosed by positive cultures and high titers of specific antibodies. Radiologically an opacity can be demonstrated in a cavity. The most reliable approach is surgical excision.

In severely compromised individuals, *A. fumigatus* or *A. flavus* may invade tissues and this is the most lethal form of the infection. Pulmonary vessels may be invaded, occluded, or even destroyed producing hemorrhage or infarction with rapid extension of the invasion through blood vessels to the brain, heart, kidneys, liver, and spleen. Where biopsy culture and serology confirm this form of the infection, treatment should not be delayed. Intravenous amphetericidin B in full dosage is usually suggested. The mortality of all invasive forms of aspergillosis, even with treatment, is high.

R. C. Vickery, M.D.; D.Sc.; Ph.D.; Blanton/Dade City, Florida.

ASPHALT (or Asphaltum). A semi-solid mixture of several hydrocarbons, probably formed because of the evaporation of the lighter and more volatile constituents. It is amorphous, of low specific gravity, 1–2, with a black or brownish-black color and pitchy luster. Notable localities for asphaltum are the Island of Trinidad and the Dead Sea region, where Lake Asphaltites were long known to the ancients. See also **Coal Tar and Derivatives; Petroleum.**

ASPHERIC SURFACE. A surface of a lens or mirror which has been changed slightly from a spherical surface as an aid in reducing aberrations. Parabolic mirrors for telescopes and the Schmidt objective are common aspheric surfaces.

ASPHYXIA. Suffocation, the consequences of interference with the aeration of the blood, usually from interference with respiration, whether by mechanical means or by the inhalation of gases containing insufficient oxygen, although it may result from other causes which would depress the respiratory center or result in a deficiency of hemoglobin in the blood. The effects are cyanosis, increased blood pressure, violent respiratory efforts, ultimately leading to unconsciousness and death if the cause is not removed. The effects are partly due to anoxemia and partly to excess of carbon dioxide in the blood. When the blood supply to a limited portion of the body is temporarily interrupted, the term *local asphyxia* sometimes is used. See also **Anoxemia.**

ASPIRIN. A drug used for nearly a century to relieve headaches and general aches and pains and to reduce the swelling and pain associated with joints (gout, ague, rheumatoid arthritis). In recent years, attention to aspirin for its apparent role in reducing heart attacks (coronary thrombosis) and strokes has increased. Trial studies also are underway for its use in reducing the risk of fatal colon cancer.

As early as 1763, the Rev. Edward Stone of Chipping-Norton, Oxfordshire (England), reported to the Royal Society that the bark of the willow tree (*Salix alba*) was found to be effective by his local constituents for treating ague. He reported his findings to the Royal Society, creating much interest. Medical historians also report that Hippocrates as well as some North American Indian tribes were aware of the analgesic effects of the bark of certain trees. Ways were sought to prepare what chemists at the time referred to as *salicin* by extracting the active ingredient from the willow bark. To produce very small amounts of salicin required several pounds of bark, causing the price to be quite high. However, with further efforts, the extraction process was improved, lowering the price. Ague and gout occurred widely, and the extract market became quite large—sufficiently large to interest early European pharmaceutical firms to find a way to synthesize the product. The Germans, who during the mid-1800s excelled in organic chemistry due to their synthesizing important dye chemicals, finally found a way to synthesize salicylic acid. Ironically, however, chemists in France were first to name the product, *l'acide salicylique* (salicylic acid).

It was found that salicylic acid in its pure form had a number of deficiencies, and for a number of years chemists sought a salicylic acid–based compound that would be effective yet less harsh and that could counteract pain with smaller dosages. This process ended in 1898 with the introduction by Bayer of acetylsalicylic acid, which has the formula $C_6H_4(COOH)CO_2CH_3$ and since then has been commonly referred to as *aspirin*.

The market for aspirin grew at a rapid rate, with sales in the United States reaching $2 billion/year in 1990. This represents 1600 tons of the drug, or 80 million tablets. Within recent years, some aspirin has been formulated with other materials. These include buffers for reducing stomach irritation experienced by some people who consume aspirin. Also within the last decade or so, other nonsteroidal anti-inflammatory drugs (NSAIDs) have been introduced into this highly competitive marketplace.

In recent years aspirin has been subjected to some negative publicity. In a minority of the population, aspirin can induce hypersensitivity syndrome. Obviously, persons who exhibit an allergic reaction to the drug are not candidates for it and should turn to other NSAIDs. Needless to say, aspirin must be used with moderation. A most unusual situation was reported by Thibault (1992) in one of the publications listed. A middle-aged man, who had a psychiatric history, complained at an emergency center of nausea, vomiting, shortness of breath, and hallucinative hearing. It was learned that the patient had consumed four aspirin tablets every 2 to 4 hours for a period of 2 weeks. Obviously, this defines *extreme immoderation* in terms of the drug's use. However, with treatment, the patient's symptoms disappeared within 2 days. Fortunately, such findings of the abuse of aspirin are rare, but the case does emphasize, in perspective, the relative safety of the drug when properly administered.

Probably the most negative situation involving aspirin arose about a decade ago, in connection with the appearance of Reye's syndrome. The biochemistry of this connection has not been fully elucidated, but based upon clinical findings, the general medical community stipulates that aspirin not be used (particularly with young children) where there any symptoms or suspicions that influenza may be present. See also **Reye's Syndrome.** Also, as a result of this incident, aspirin was removed from the World Health Organization's list of essential drugs, representing a decision that was not universally accepted by medical professionals.

Biochemistry of Aspirin: The biochemical paths and actions by which aspirin and other salicylates achieve their therapeutic effects were poorly understood until at least a partial mechanism was proposed by Sir John Vane in 1971. Vane, who later received a Nobel Prize for his efforts (1982), found that NSAIDs, including aspirin, block the production of prostaglandins by cells and tissues. During the same time frame, Vane and other researchers also confirmed the inhibitory effects of aspirin on platelet aggregation, this caused by interference with the ability of platelets to synthesize prostaglandins, notably thromboxane A_2. The complexities of the topic go well beyond the scope of this volume, but are well ventilated in the Vane (1971), the Smith-Willis (1971), and the Weissmann (1991) articles listed. See also **Prostaglandins**.

Much current research relating to aspirin and heart attacks and strokes is going forward, principally in the form of trial study groups, with emphasis on the effects of dosage. The findings of aspirin's advantage in connection with fatal colon cancer are in their early and debatable study phases.

Additional Reading

Abramson, S., et al.: "Modes of Action of Aspirin-Like Drugs," *Proceedings, National Academy of Sciences* 82(21) 7227 (November 1985).

Bashein, G., et al.: "Preoperative Aspirin Therapy and Reoperation for Bleeding After Coronary Artery Bypass Surgery," *Arch. Intern. Med.*, 114, 835–9 (1991).

Dutch TIA Trial Study Group: "A Comparison of Two Doses of Aspirin (30 mg vs. 283 mg a day) in Patients After a Transient Ischemic Attack or Minor Ischemic Stroke," *N. Eng. J. Med.*, 1261 (May 7, 1992).

Ferreira, S. H., and J. R. Vane: *Annual Review of Pharmacology*, Vol. 14, 57 (1974).

Mills, J. D.: "Aspirin, The Ageless Remedy?" *N. Eng. J. Med.*, 1303 (October 31, 1991).

Pederson, A. K., and G. A. FitzGerald: "Dose-Related Kinetics of Aspirin: Presystemic Acetylation of Platelet Cyclooxygenase," *N. Eng. J. Med.*, 1206 (November 6, 1984).

Smith, J. B., and A. L. Willis: "Aspirin Selectivity Inhibits Prostaglandin Production in Human Platelets," *Nature-New Biology*, 231, 235 (1971).

Thibault, G. E.: "The Landlady Confirms the Diagosis (Aspirin Overdose)," *N. Eng. J. Med.*, 1272 (May 7, 1992).

Vane, J. R.: "Inhibition of Prostaglandin Synthesis As a Mechanism of Action for Aspirin-Like Drugs," *Nature-New Biology*, 231(25) 232 (June 23, 1971).

Weissmann, G.: "Aspirin," *Sci. Amer.*, 84 (January 1991).

ASSASSIN BUG (*Insecta, Hemiptera*). Any bug of the large predacious species constituting the family *Reduviidae*. The assassin bug is found in the southern part of the United States and in the West Indies. Throughout the world, it is estimated that there are about 2500 species. Some species pounce upon their prey; other species stick their legs into resin from a tree and hold the sticky limbs aloft, awaiting a likely victim to come along. Some assassin bugs secrete a fluid which other insects find attractive, but the fluid has an intoxicating effect on likely victims, thus making them easy prey. An oily hair on the legs of the assassin bug helps in holding prey. The thorax of the assassin bug also produces a poisonous venom which, when injected into prey, assists in reducing the tissues of the victim to a thick juice ready for convenient consumption and assimilation.

ASSEMBLER (Computer System). A computer program which operates on symbolic input data to produce machine instructions by carrying out such functions as (1) translation of symbolic operation codes into computer instructions, (2) assigning locations in storage for successive instructions, or (3) assignment of absolute addresses for symbolic addresses. An assembler generally translates input symbolic codes into machine instructions item for item, and produces as output the same number of instructions or constants which were defined in the input symbolic codes.

Assembler language may be defined as computer language characterized by a one-to-one relationship between the statements written by the programmer and the actual machine instructions performed. The programmer thus has direct control over the efficiency and speed of the program. Usually, the language allows the use of mnemonic names instead of numerical values for the operation codes of the instructions and similarly allows the user to assign symbolic names to the locations of the instructions and data. For the first feature, the assembler contains a table of the permissible mnemonic names and their numerical equivalents. For the second feature, the assembler builds such a table on a first pass through the program statements. Then, the table is used to replace the symbolic names by their numerical values on a second pass through the program. Usually, some dummy operation codes (or pseudocodes) are needed by the assembler to pass control information to it. As an example, an origin statement is usually required as the first statement in the program. This gives the numerical value of the desired location of the first instruction or piece of data so that the assembler can, by counting the instructions and data, assign numerical values for their symbolic names.

The format of the program statements is usually rigidly specified and only one statement per input record to the assembler is permitted. A representative statement is: symbolic name, operation code (or pseudocode), modifiers and/or register addresses, symbolic name of data. The mnemonic names used for the operation codes usually are defined uniquely for a particular computer type with little standardization between computer manufacturers even for the most common operations. The programmer must learn a new language for each new machine with which he works.

An example of a program prepared in an assembler language follows. The explanatory comments following the REM (remarks) mnemonic and those to the right of the other program statements are ignored by the assembler program and thus do not affect execution of the program. See also **Language (Computer).**

```
       ABS
       ORG    100
       REM    THIS IS AN EXAMPLE OF A PRO-
       REM    GRAM IN AN ASSEMBLER LAN-
       REM    GUAGE. "ABS", "ORG", "CALL",
       REM    "REM", "DC", "DEC" AND "END"
       REM    ARE PSEUDOCODES. THE OTHER
       REM    MNEMONICS IN THE SAME COL-
       REM    UMN ARE OPERATION CODES.
       REM    THE NAMES SUCH AS A, B,
       REM    START, AND OVER ARE SYM-
       REM    BOLIC ADDRESSES ASSIGNED BY
       REM    THE PROGRAMMER

START  SLT    32     THESE INSTRUCTIONS
       LD     A      COMPUTE A/B*C AND
       D      B      PUT THE VALUE IN D
       M      C
       SLT    1
       STO    D
       CMP    D4B1   THESE INSTRUCTIONS
       BSC    Z+     COMPARE D WITH 4 AND
       MDX    OVER   3.2 AND BRANCH OUT IF
       CMP    D32B1  IT DOES NOT LIE BE-
       BSC    Z-     TWEEN THEM
       MDX    UNDER

EXIT   CALL   EXIT   THE ASSEMBLER SETS
       REM           UP A BRANCH TO THE
       REM           SYSTEM EXIT SUBROU-
       REM           TINE (NAMED "EXIT")

OVER   LD     D1     THESE INSTRUCTIONS
       STO    EROR   SET THE VALUE OF
       MDX    EXIT   EROR TO 1 BEFORE
       REM           EXITING

UNDER  SRA    16     THESE INSTRUCTIONS
       STO    EROR   SET THE VALUE OF
       MDX    EXIT   EROR TO ZERO BEFORE
       REM           EXITING

       ORG    200    THIS SET OF PSEUDO-
A      DC     1      CODES IS USED BY THE
B      DC     2      ASSEMBLER TO SET UP
C      DC     3      THE DATA AND DATA
D      DC     0      WORKSPACE FOR THE
EROR   DEC    -1.0   ABOVE PROGRAM
D1     DEC    1.0
D4B1   DEC    4.0B1
D32B1  DEC    3.2B1
       REM           LAST STATEMENT IN
       REM           PROGRAM
       END
```

Thomas J. Harrison, International Business Machines Corporation, Boca Raton, Florida.

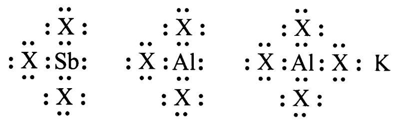
ASSOCIATION (Chemical). The combination of molecules of the same substance to form larger aggregates consisting of two or more molecules. See also **Elastomers;** and **Molecule.**

Association was first thought of as a reversible reaction between like molecules that distinguished it from polymerization, which is not reversible. Association is characterized by reversibility or ease of dis-association, low energy of formation (usually about 5 and not more than 10 kcal per mole), and the coordinate covalent bond which Lewis called the acid-base bond. Association takes place between like and unlike species. The most common type of this phenomenon is hydrogen bonding. Association of like species is demonstrable by one or more of the several molecular weight methods. Association between unlike species is demonstrable by deviation of the system from Raoult's law.

The strength of the coordinate covalent bond is a function of polarity of the associating molecules. Hence, associated molecules vary in stability from very unstable to very stable. The argon-boron trifluoride complex is quite unstable, whereas calcium sulfate dihydrate (gypsum) is very stable. The bond strength associated with stability has been measured for a number of combinations. The strengths of some hydrogen bond types decrease in the order FHF, OHO, OHN, NHN, CHO, but they are dependent upon the geometry of the combination and upon the acid-base characteristics of the group. Steric effects can have a marked effect on the strength of the coordinate covalent bond. This was demonstrated by a study of the strength of the series NH_3, $C_2H_5NH_2$, $(C_2H_5)_2NH$, $(C_2H_5)_3N$ as bases toward an acid in solution and in the gaseous state and the comparison of the base strength of triethylamine and quinuclidine. The latter is, in effect, triethylamine in which the two-carbon atoms of each ethyl group are tied together by another carbon. The geometry of the ethyls around the nitrogen is drastically changed, and the cyclic is a stronger base than the triethyl compound. The factors affecting the strength of the hydrogen bond also influence the degree of association.

Association within the same species accounts for the high boiling points of water, ammonia, hydrogen fluoride, alcohols, amines, and amides. Ethyl ether and butanol contain the same number of atoms of each element, but butanol has a boiling point of 83°C above that of ethyl ether as a result of more extensive hydrogen bonding. Some substances associate completely to two or more formula weights per molecule. Carboxylic acid, by a hydrogen-to-oxygen association, form dimers with a six-membered ring. N-unsaturated amides dimerize in the same manner, whereas N-substituted amides dimerize in a chain form in a *trans* configuration.

Hydrogen bonding is so common that coordinate bonds between other elements are sometimes overlooked. Antimony(III) halides form very few complexes with other halides, whereas aluminum halides readily form complexes. The octet of electrons is complete in all atoms of the antimony halides, but is incomplete in the aluminum atom of aluminum halides:



(a) (b) (c)

Aluminum can accept two electrons to complete its octet. The pair of electrons is available from the halogen. An alkali halide can supply the electrons and form a complex (c), or the electron pair may come from the halogen of another aluminum chloride. Association with other aluminum halides accounts for the higher melting point of aluminum halides over antimony(III) halides which have a formula weight of 95 or more. The association of aluminum sulfate, alkali metal sulfate, and water to form the stable alums is one of the more complex examples.

The formation of solvates is association between unlike species. Solvation is more frequent between substances of high polarity than those of low polarity. This is illustrated by the decrease in the tendency to form solvates with decrease in dipole moment and dielectric constant (shown in parentheses) for N-methylacetamide (3.59; 172), to water (1.84; 78.4), to ethanol (1.70; 24.6); to ammonia (1.48; 78.4); to ethanol (1.70; 24.6); to ammonia (1.48; 17.8); to methylcyclohexane (0; 2.02) for which few associations are known.

J. A. Riddick, Baton Rouge, Louisiana.

ASSOCIATION (Coefficient of). In statistical theory, this word is used (a) in a general sense, to denote the degree of dependence between two variables; and (b) especially to denote the relationship between two variables which are simply dichotomized. For example, if a set of *n* numbers is classified as A or not $-A$, and as B or not $-B$, an association table is of the following kind:

	A	Not $-A$	Totals
B	a	b	$a + b$
not $- B$	c	d	$c + d$
Totals	$a + c$	$b + d$	$a + b + c + d = n$

A coefficient of association Q is defined by

$$Q = \frac{ad - bc}{ad + bc}$$

It can vary from -1 to $+1$ according to the strength of the association. Other coefficients are sometimes used, especially

$$V = \frac{ad - bc}{\{(a + c)(b + d)(a + b)(c + d)\}^{1/2}}$$

The latter is, in fact, related to chi-square for the fourfold table as $V = \{\chi^2/n\}^{1/2}$.

Sir Maurice Kendall, International Statistical Institute, London.

ASSOCIATION (Ecology). Central to certain concepts of ecology is the interaction between various otherwise unrelated species in a way that is beneficial to the participating parties, but not always indispensable. Ants and plant lice are sometimes associated in this way. The plant lice are guarded by the ants and sometimes carried to a good food supply whereupon the ants receive the sweet honey dew secreted by their charges. So-called cleaner fishes play useful roles in removing barnacles and other deposits from larger fishes and in recognition for these services are not eaten by the larger fishes. This type of association is also sometimes termed *commensalism*.

Plants. It is well established that plants are not distributed in nature in a haphazard fashion, but in habitats in which when certain species are present certain others usually occur also. Each such community of plants, composed of more or less the same group of species, is called a plant association. Some plant associations, such for example as the marginal rush or cattail association around a pond, may occupy only localized areas. Other associations, such as some of the grassland or desert shrub associations of western North America may occupy vast continuous areas. In general, however, plant associations are the smaller units of vegetation occurring within a plant formation (see below) and their distribution is largely controlled by local soil and climatic conditions. Local differences in climate, in turn, are largely a function of topography. Some plant associations, such as a lichen association on a rock cliff, are relatively simple in organization. Others are relatively complex. The oak-hickory association of the eastern United States, for example, is named for the two prominent genera of trees present. Associated with the oaks and hickories, however, are occasional other large trees. In addition there are usually present smaller kinds of trees, species which constitute a shrub layer, and herbaceous species which constitute a more or less continuous ground cover.

A larger unit of vegetation than the association is the formation. A *plant formation* usually occupies very large regions and its limits are controlled primarily by climatic conditions. Some of the major plant formations of North America are the tundra, the boreal forest, the hemlock-hardwood forest, the deciduous forest, the grasslands, the western coastal forest, the western mountain forest, the semi-deserts, and the tropical forests. Within each formation there are usually many different plant associations. Most plant associations are not permanent, but in the phenomenon of plant succession one association gradually replaces another. Many successions are in progress in any plant formation. The end results of the successional replacement of one plant association in turn by another is, if the process goes to completion, the establishment of a *climax association*. Such an association is a stable plant community and is not succeeded by any other association; it is the apex of the successional process and, barring changes of climate, will continue to reproduce itself indefinitely.

Animals. While most animals are solitary, associating with others of their kind only incidentally or during the breeding season, others normally live in some relationship with members of the same or of other species.

The simplest association of members of the same species is gregariousness. Gregarious animals are not bound by the association but profit by it. Examples are the great herds of herbivorous animals such as the bison and the packs of predacious animals, such as wolves.

Colonial association may be accompanied by structural union between individuals, as in many marine polyps, or may be based on behavior, as in the social insects. The term merges with social organization. This type of association is accompanied by structural specialization of individuals for special tasks, except in human society where it depends on specialized training.

The association of individuals of different species may be the relatively loose type called commensalism in which both forms benefit but not in an essential way, or the indispensable symbiosis in which neither organism can persist without the other. An excellent example of symbiosis is the relation of termites with the protozoa found in their intestine; neither can live without the other.

An association in which one individual lives at the expense of the other is called parasitism.

Slavery is an association practiced by some of the social insects and, at one time, by people; among the insects the slaves are of a different species.

Such relations as symbiosis and parasitism also occur among plants, where they are exemplified by the combining of algae and fungi to form lichens and by the mistletoe which is parasitic on trees. Symbiotic relations between animals and plants also occur.

ASTATIC. A term meaning without orientation or directional characteristics.

ASTATINE. Chemical element symbol At, at. no. 85, at. wt. 210 (mass number of the most stable isotope), periodic table group 17, classed in the periodic system as a halogen, mp 302°C, bp 337°C. All isotopes are radioactive. This element occurs in nature only in minute amounts as a result of minor branching in the naturally occurring alpha decay series: $^{218}\text{At}(t_{1/2} = \text{ca. } 2 \text{ sec.})$ is produced to the extent of 0.03% by the beta decay of ^{218}Po(radium A), 99.97% going by alpha decay to ^{214}Pb(RaB); $^{216}\text{At}(t_{1/2} = 3 \times 10^{-4} \text{ sec.})$ 0.013% by beta decay from ^{216}Po(thorium A);

$$^{215}\text{At}(t_{1/2} = 0.018 \text{ sec.})$$

0.0005% by beta decay from ^{215}Po(actinium A). Astatine-217 ($t_{1/2} = 0.020$ sec.) is a principal member of the neptunium ($4n + 1$) series, all members of which occur only to that extent to which the parent ^{237}Np is produced by naturally occurring slow neutrons from uranium.

The first isotope to be discovered was ^{211}At made by Carson, Mackenzie, and Segrè by bombardment of a bismuth target with α-particles from the 60-inch cyclotron at Berkeley in 1940. The reaction is $^{209}\text{Bi}(\alpha, 2n)\,^{211}\text{At}$. The half-life of ^{211}At is 7.2 hr. It decays in two modes, 60% by K-electron capture and 40% by α-particle emission. The longest-lived isotope is $^{210}\text{At}(t_{1/2} = 8.3 \text{ hr.})$; other isotopes having half-lives longer than 1 hr are 206, 207, 208, and 209. Various of the collateral radioactive series involving bombardment reactions contain other astatine isotopes, such as ^{214}At and ^{216}At. All these isotopes have half-lives that are only fractions of a second. The total number of isotopes is at least nineteen, including spallation reaction products as well as bombardment ones. They also include two short-lived isotopes, ^{215}At and ^{218}At, occurring in very small amounts in the branched β-disintegration of ^{215}Po(actinium A) and ^{218}Po(radium A), respectively, as noted above.

The chemistry of astatine determined by tracer techniques, is in keeping with the regular transition of properties of the halogens. The acid properties of astatine are less marked than those of iodine, while its electropositive character is more marked than that of iodine. After reduction by SO_2 or metallic zinc, the astatine activity is carried by silver iodide or thallium iodide, so it evidently forms insoluble silver and thallium salts. This represents astatine in the univalent negative state characteristic of the halogens. However, astatine is very readily oxidized by bromine and ferric ions, giving indications of two higher oxidation states. Although there is no evidence from migration experiments of the presence of positive ions in the solution, astatine deposits on the cathode, as well as on the anode, in the electrolysis of oxidized solutions. Elemental astatine can be volatilized, although not so readily as iodine, and it has a specific affinity for metallic silver. The similarity to iodine is also shown by the observation that astatine concentrates in the thyroid glands of animals.

Additional Reading

Fisk, Z., et al.: "Heavy-Electron Metals: New Highly Correlated States of Matter," *Science*, 33 (January 1, 1988).

Ghiorso, A., et al.: "Preparation of Transplutonium Isotopes by Neutron Irradiation," *Phys. Rev.*, 78(4) 472 (1950).

Hammond, C. R.: "The Elements," in *Handbook of Chemistry and Physics*, 67th Edition, CRC Press, Boca Raton, Florida, 1986-1987.

Hyde, E. K.: "Astatine," in *McGraw-Hill Encyclopedia of Chemistry*, McGraw-Hill, New York, 1983.

Kent, J. A.: *Riegel's Handbook of Industrial Chemistry*, Van Nostrand Reinhold, New York, 1983.

Staff: *Handbook of Chemistry and Physics*, 73rd Edition, CRC Press, Boca Raton, Florida (1992–1993).

ASTERISM. One of the characteristic effects sometimes observed in x-ray spectrograms. It has, roughly, the shape of a star, and commonly indicates the presence of internal stress in the material under investigation.

ASTEROID. During the 19th century and much of the 20th century, asteroids were considered to be made up of the debris that resulted from the shattering of one or more ancient planets, giving rise to the term *minor planets.* This earlier theory, however, did not postulate what may have caused one or more ancient planets to shatter. The most recent and widely accepted theory considers the asteroids to be remnants of early planetary material that failed to coalesce into a planetary body. The newer theory presents a major difference in the perspective of asteroid scholars.

The first few asteroids to be discovered (1801–1807) were considered to be minor planets. These are the larger of the asteroids known today—*Ceres, Pallas, Juno,* and *Vesta.* Today, the asteroid population is estimated in the millions, ranging widely in mass and dimension. Even then, collisions among asteroids of any appreciable size are rare because of the immense emptiness that exists between planets of the solar system.

Search for the "Missing Mass": As early as 1766, Titius, a German astronomer, studied the relative spacing of the planets and from his calculations observed a "missing planet" that should be found between the orbits of Mars and Jupiter. This excited European astronomers to search for the missing planet, and, as previously mentioned, Ceres, Pallas, Juno, and Vesta were observed by 1807.

Also regarded as minor planets by Herschel, he suggested that they be given a special group name, *asteroids*, from the Greek "starlike," because they appeared through the telescope more like distant stars than distant planets, such as Uranus, which had been discovered by Herschel in that time period.

Early Discoveries of Asteroids. The first asteroid discovered, Ceres, was found accidentally by Piazzi on January 1, 1801. His attention was directed to it by noticing the motion of the object through the stars. As the object approached the position of the sun, there was danger of its being lost, for the methods of orbit computation were not well developed at that time. The mathematician Gauss went to work on the problem and invented his well-known method for orbit computation, by means of which he was able to predict positions permitting the rediscovery of Ceres after it has passed the sun. Since the orbit was found to lie in the gap between the orbits of Mars and Jupiter, and the object was found to have a mean distance from the sun of 2.8 astronomical units, strong support was given by it to Bode's relation. See also **Bode's Relation**.

Continuing Discoveries. Up to the middle of the nineteenth century, only five more asteroids were discovered; but with the application of photography to astronomy, the discoveries became more and more frequent, until, at the present time, many hundreds are under observation. They are first detected by noticing the movement of a starlike object through the stars. Photographically, if the camera is arranged to follow the motions of the stars, the star images will appear as dots on the plate, while the asteroid image will be trailed out into a short line. The most extensive program of search for asteroids was carried on by Wolf at Heidelberg during the period following 1891. From this time on through the first two decades of the present century. Wolf and his assistants are credited with no less than 500 discoveries. Hundreds of asteroids are now picked up each year in the course of other investigations.

Rendezvous with Galileo

Unless the satellites of Mars (*Deimos* and *Photos*) prove to be asteroids, the first close-up of an asteroid occurred on October 29, 1991, when the NASA spacecraft *Galileo* returned an image of *Gaspra*, a comparatively small (12 × 20 km) asteroid that orbits between Mars and Jupiter. *Gaspra*, a stony S-type asteroid, is oddly shaped, possibly the result of numerous collisions. One reseacher estimates *Gaspra*'s age at between 300 and 500 million years, which makes it comparatively young on the basis of the solar system, which is estimated at 4.6 billion years. The spacecraft's ability to transmit additional images will depend upon remotely fixing an antenna. Originally, a resolution of features less than 100 meters across had been planned.

Designating Asteroids. When an asteroid is first discovered, it is designated first by the year of discovery, then by two letters that indicate the half of the month in which the object was found, and last by the chronological order of discoveries within that half-month. After the orbit of the object has been determined, and if it proves to be a new asteroid, it is assigned a permanent number, in chronological order of discovery, and the discoverer is privileged to name the object as he may choose. In general, asteroids are given Latinized names with feminine endings.

Physical Characteristics of Asteroids. In the accompanying table are summarized key physical parameters of the larger and better-known asteroids. It is well known that the reflected sunlight from many of these objects varies in a periodic manner that can be explained adequately only on the basis of rotation of the object. In the case of Eunomia, it has been shown definitely that the object must be close to spherical and that the variation in light is due to different reflecting powers of different parts of the surface. On the other hand, Eros has been shown to have a brick-like shape, with the light variations due to rotation of this irregular object. Several techniques in recent years have contributed to improved asteroid imagery, including radar observations. In terms of estimating total size of an asteroid, thermal (infrared) methods are used. On the average it appears that surface temperatures of asteroids are ap-

REPRESENTATIVE ASTEROIDS

Asteroid	Diam. (km)	Mass(10^{15}g)	Period (d)	*a*(A.U.)	*e*	*i*(deg.)
Ceres	933	60×10^7	1681	2.767	.08	10.6
Pallas	523	18×10^7	1684	2.767	.24	34.8
Juno	220	2×10^7	1594	2.670	.26	13.0
Vesta	501	10×10^7	1325	2.361	.09	7.1
Hebe	220	20×10^6	1380	2.426	.20	14.8
Iris	200	15×10^6	1344	2.385	.23	5.5
Hygiea	320	60×10^6	2042	3.151	.10	3.8
Eunomia	280	40×10^6	1569	2.645	.18	11.8
Psyche	280	40×10^6	1826	2.923	.14	3.1
Nemausa	80	9×10^5	1330	2.366	.06	9.9
Eros	14	5×10^3	642	1.458	.22	10.8
Davida	260	3×10^7	2072	3.182	.18	15.7
Icarus	1.4	5	408	1.077	.83	23.0
Geographos	3	50	507	1.244	.34	13.3

NOTE: In the main asteroid belt between Mars and Jupiter, there are approximately 1000 asteroids that are larger than 30 km in their longest direction. Of these, about 200 are larger than 100 km across. Researchers estimate from calculations, that there must be a million or more asteroids having a longest direction of 1 + km. With numerous asteroids, the term *diameter* is inappropriate because of their odd, nonspherical shape.

proximately 200 kelvins. This, of course, varies with the albedo of the object and its diameter and distance from the sun. An interesting laboratory technique that can be used to model asteroid shapes is described by Binzel, et al. See Additional Reading. This system creates synthetic light curves for varying object shapes.

Asteroid Locations (Belts)

The greater number of asteroids observed and cataloged lie in what is commonly called the "main belt," located between Mars and Jupiter.

The orbits of the asteroids have been studied carefully ever since the discovery of Ceres. See Fig. 1. In fact, this group of objects may be considered as a laboratory in which the workers in the field of celestial mechanics may test out various theories. Because their orbits lie between the orbits of Mars and Jupiter, and because their masses are very small, the asteroids have large perturbations exerted upon them by the planets; whereas the planets themselves are virtually unaffected by the asteroid attractions. Many of the methods of computing perturbations were developed as the result of research on the orbits of minor planets. One particularly interesting result is found in the case of the Trojan asteroids.

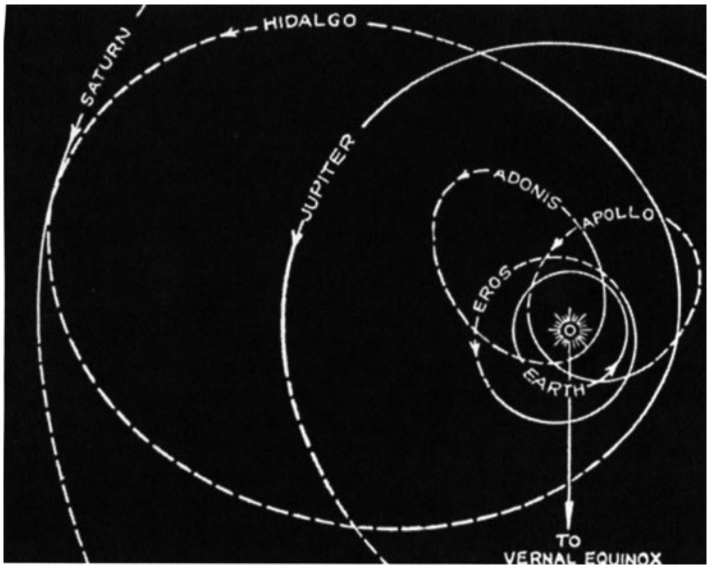


Fig. 1. Orbits of four unusual asteroids. Broken lines represent parts of the orbits south of the plane of the ecliptic.

Amor Asteroids. A group of asteroids with orbits that cross the orbit of Mars (as projected on the ecliptic plane), but do not cross the orbit of the earth. Typical Amor orbits reach from the asteroid belt to a point between the earth and Mars. The group is named after the prototype asteroid, Amor.

Apollo Asteroids. A group of asteroids with orbits that cross the orbit of the earth (as projected on the ecliptic plane). Collisions with the earth are possible and apparently have occurred as described later, but these asteroids generally cross above or below the ecliptic plane, thus minimizing the possibilities of collisions. A substantial number of Apollo asteroids have been observed and their orbits have been calculated. In other instances, some have been discovered and ultimately "lost." Instrumental evidence suggests that they are rocky bodies, generally a few kilometers across. It is believed that most meteorites may be fragments broken off from them.

Trojan Asteroids. Two groups of asteroids in Jupiter's orbit about 60° ahead of and behind Jupiter. Existence of such bodies was predicted by Lagrange. Each member of the Trojan group has a period and mean distance nearly identical to those of the planet Jupiter. The group is of considerable theoretical interest because it represents examples of the solution of the three-body problem proposed by Lagrange, who proved theoretically that an object so located that it is equidistant from both Jupiter and the sun would be in a stable position, i.e., would remain there and continue to go about the sun with the same period as Jupiter. The members of the Trojan group all behave in approximately this manner, each of them being 20° of the vertex of an equilateral triangle, with the sun and Jupiter at the other vertices. They move about this vertex in a complex curve, and will remain in this vicinity unless they are perturbed by the attraction of Saturn.

Material Composition of Asteroids

Spectroscopic measurements yield information on the chemical composition of asteroids. There are several letter-designated classes in terms of materials. This taxonomic system developed in 1984 has since been refined by astronomers at the University of Hawaii. The system does not commence with S or proceed alphabetically in a logical manner, but rather letters were assigned in order of their observed relative abundancies and thus is confusing to the uninitiated. To achieve some simplification, asteroids currently are grouped into classes by composition, namely, the *primitive* asteroids, which include types C, D, and P; the *igneous* asteroids, which encompass the S, M, and E types; and the *metamorphic* asteroids, which contain the F, G, B, and T types.

Primitive asteroids are found in the outlying portions of the main belt. Theoretically, it is suggested that the primordial process which produced them was one of such large large temperature gradient that the composition of the asteroids essentially was altered at the time of their origin. Being farthest from the sun, these asteroids are assumed to be rich in carbon and water. Today, they are considered to be representative of the material left over from when the solar system was formed.

Igneous asteroids are found closest to the sun and must have endured severe heating. It is assumed that they formed complex mineral mixtures during melting and solidification. Some researchers suggest that radioactive heating may have been involved. The lack of similarity of materials composition among the asteroids also has been attributed by some researchers as "space weathering."

Metamorphic asteroids are found in the central region of the belt.

The composition similarities between asteroids and meteorites is mentioned briefly under **Meteorides and Meteorites**.

Similarity with Comets. When observed at great distances, comets have been mistakingly identified as asteroids and vice versa. This proved to be the case of *Chiron*, at first considered an asteroid after its discovery in 1987. *Chiron* was found to have unusual characteristics, including its distant orbit between Saturn and Uranus. Observers now generally agree that *Chiron* is a comet, with the requisite properties required of a comet.

Hirayma Asteroid Families usually are classified as C, S, and M types. They have similar orbits and are suspected to be fragments from collisons between pairs of asteroids. They were observed early in the 20th century (by Hirayama, 1918–1929). The use of reflection spectroscopy, polarimetry, and thermal radiometry have provided information on the composition of these minor planetary bodies, suggesting new insights as to the importance of the Hirayama bodies. If these asteroids are the result of collisons, then the members of a single family should show identical composition, assuming that the parent body was homogeneous. But, on the other hand, if the dynamical families were formed by the collisional focusing of unrelated field asteroids, then one would expect to find only the pattern of compositions that is characteristic of that region of the asteroid belt. A technique known as UBV photometery (U-B ultraviolet-minus-blue and the B-V blue-minus-visual) has shown that the colors of minor planets indicate compositions quite distinct from those of the field population in each of the three Hirayama families. Researchers have observed that the Eos and Koronis families apparently originated from the collisonal fragmentation of undifferentiated silicate bodies and the Nysa group from a geochemically differentiated parent body. Considerable research remains to further refine the postulations concerning the origin of the Hirayama families.

Captured Asteroids

The irregular shapes and other unexpected characteristics of Deimos and Phobos, the moons of Mars, as revealed by high-resolution photos obtained by the Viking orbiters, have suggested to a number of specialists that these two bodies are asteroidal satellites. If so, they are the first captured asteroids to be viewed at close range. See Figs. 2 and 3. As of the early 1990s, this postulation has been neither proven or disproven.

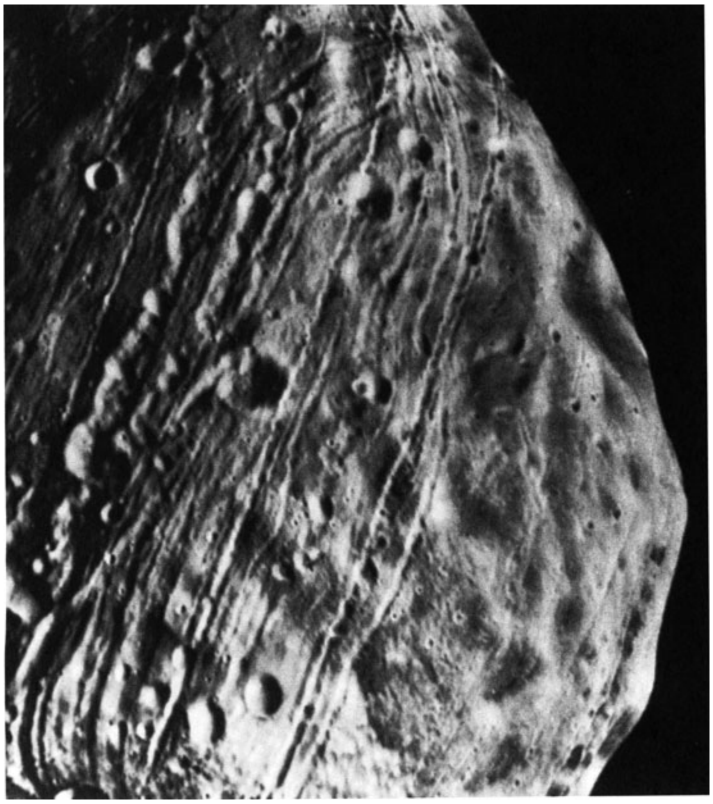


Fig. 2. Phobos, a satellite of Mars, may be a captured asteroid. (*Viking Orbiter 1.*)

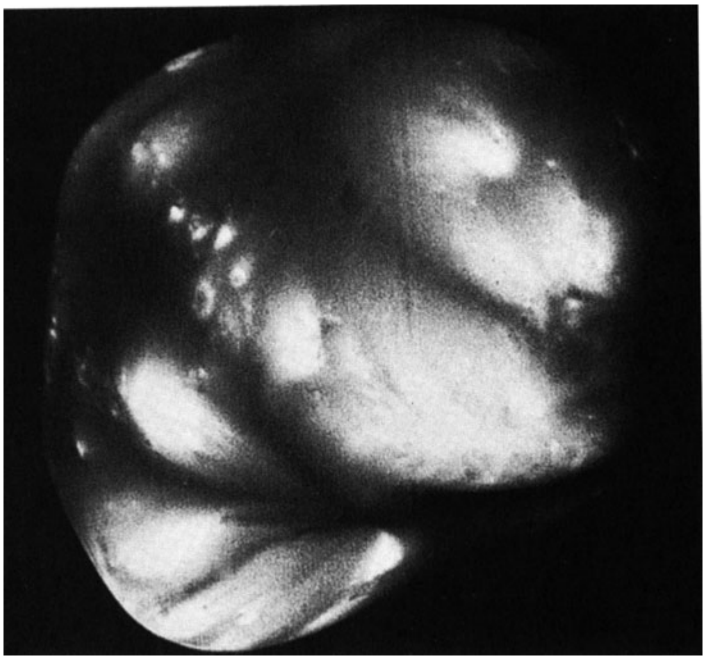


Fig. 3. Deimos, a satellite of Mars, in nearly full phase. This moon may be a captured asteroid. (*Viking Orbiter 2.*)

Binary Asteroids

Light curves obtained of 624 Hektor (Fig. 4) have suggested that the asteroid has the shape of a dumbbell or possibly a fat cigar, considerably longer than it is wide. The irregular form of an asteroid like this may result from collisions with other asteroids, but, on the other hand, it is unlikely that a collision would produce an oblong body. Other observers have suggested that Hektor may not be a dumbbell as previously proposed, but rather two asteroids in contact orbiting together. Future spectroscopic observations may reveal whether Hektor is one or two objects.

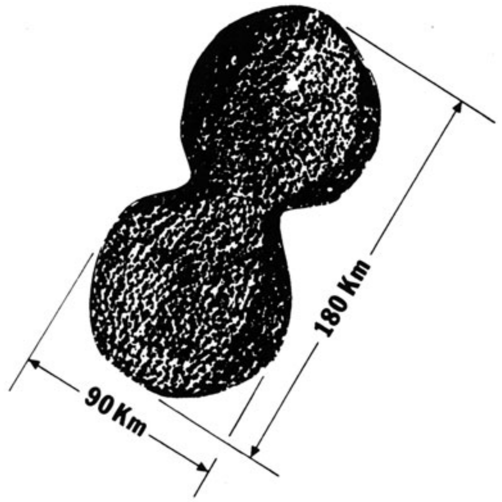


Fig. 4. Light curves indicate that Asteroid 624 Hektor may be a dumbbell shape, as shown here, or it may be two separate bodies (a binary asteroid).

Star Occulted by Asteroid

In a rare event which occurred on June 7, 1978, a star was eclipsed by the asteroid 532 Herculina. While the total occultation lasted for 20.6 seconds, there were six additional dimunitions in starlight within 2 minutes of the main eclipse. These ranged from 0.5 to 4 seconds. Since the star was almost four magnitudes, atmospheric interference was ruled out. The longest secondary occultation has been confirmed by an independent observation. This occultation was caused by a secondary body about 50 kilometers in diameter and about 1000 kilometers from Herculina, whose diameter is 220 kilometers. Current thinking is that the six secondary eclipses were caused by six satellites of Herculina. The masses and distances are such that the system would be gravitationally stable.

Asteroid Collisions with Earth

To some specialists in the field, it is not a question of will an asteroid collide with the earth, but of the probability of such an event. There is growing evidence of past encounters of this nature. As shown by the map of Fig. 5, craters on earth are relatively numerous. The geographical regions of central and eastern Canada embrace nearly 50%

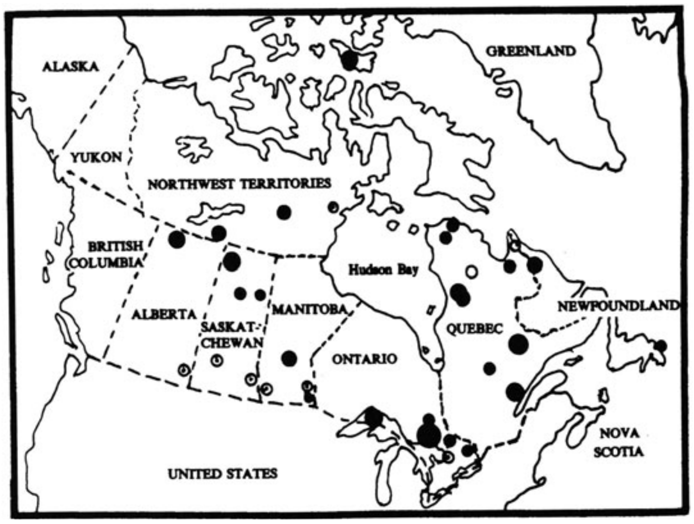


Fig. 5. Concentration of ancient (asteroid) craters in Canada, particularly in central and eastern portions. Over twenty craters (black circles) have been identified. Circles with dots identify locations of possible impact structures. These craters range in age from an estimated 1.8 billion years to less than 5 million years. The oldest crater is located near Sudbury, Ontario; the youngest crater is located in extreme northern Quebec, east of Hudson Bay. Size of circle approximates size of crater (as compared with other craters). (*Drawn from data provided by Earth Physics Branch, Department of Energy, Mines and Resources, Canada.*)

of all known ancient impact craters although the region represents only about 1% of the earth's land surface. Evidence of craters also has been found in Europe, Asia, and other areas. However, there are no known regions with the similar concentration of that in Canada. This has not been satisfactorily explained to date except in a rather qualitative way—to the effect that the Canadian scientists have conducted a much more extensive survey for craters. There is also the likely possibility that many craters are simply undiscovered in other areas because they are covered over with the debris of millions of years in the past and may be located in regions that are not particularly active in terms of other geologic interests. In commenting on the Canadian craters, as well as craters elsewhere, one scientist has observed that, on the basis of the best available crater count, it is estimated that during the past 600 million years about 1500 Apollo objects (about 1 kilometer in diameter) or larger have struck the earth. This estimate assumes that about 70% of these bodies fell into the sea rather than on land. See also **Astrobleme.**

The relatively few craters on the earth and those many more observed on our moon and on many of the satellites of other planets are believed to have been caused by asteroid and/or comet impacts, among meteorites and other causal factors. The investigations of such impacts and their effects upon the earth not only fall within the realm of astrophysics, but of paleontology and paleogeology as well. Over the past two million years, there have been five relatively rapid environmental changes which have affected the biomass of the earth. There are hypotheses which attempt to explain periods of glaciation and intervening periods of warmer climates. That all or some of these environmental changes have occurred as the result of extraterrestrial forces has been discussed for many years. Shortly after the discovery of Ceres in 1801, there were proposals that asteroids colliding with the earth have been the principal cause of the major environmental periods, and, in particular, the most severe of these changes, which occurred at the end of the Cretaceous period and beginning of the Tertiary period, about 65 million years ago, when life epitomized by the dinosaurs became extinct. Paleontologists for several decades generally have not considered this hypothesis seriously and for understandable reasons, have opted for the more gradual causes of the Cretaceous-Tertiary extinctions, as contrasted with a single catastrophic event. The gradualists base these opinions in part upon the lack of geological and fossil evidence that would support the asteroid hypothesis. Further, no impact of an asteroid has been recorded during the time of recorded history.

As early as 1973, Urey speculated on the impact of a comet as the event which ended the Cretaceous period. The nuclei of comets are estimated within the same size range as an acceptable value for an impacting asteroid, i.e., from 1 to 10 kilometers (0.6-16 miles) in diameter. It is further observed, however, that comets, unlike asteroids, are composed of much ice and other substances that tend to reduce the comet size during swings close to the sun which volatilize these materials and further enrich a nebulous coma. See also **Comet.** If the estimate of the size of comets is relatively reliable, authorities suggest that observable comets have not been sufficiently abundant to produce the number of large craters on the moon, but that the abundance of Apollo objects has been sufficient to cause these craters. It has been suggested that perhaps comet impacts account for up to 35% of the larger lunar (and possibly a few earth craters), whereas the other craters have been caused by impacting asteroids and meteoroids.

The number of asteroids crossing the earth's orbit of a size sufficient to cause the five major environmental events in the earth's history (one possible example, the end of the Cretaceous period) would have required a hit about once every 100 million years. Some authorities believe that abundance of Apollo objects is sufficient to cause four collisions per every million years. Within the last few years, Alvarez and Alvarez (University of California at Berkeley) and Asaro and Michel (Lawrence Berkeley Laboratory) have located direct physical evidence (as contrasted with biological changes seen in the paleontological record) for an unusual event at exactly the time of the extinctions in the planktonic realm. A hypothesis has been developed to explain nearly all the available paleontological and physical evidence (Alvarez, et al, 1980). The Cretaceous-Tertiary boundary layer has been inspected in a number of locations, including Denmark and Italy. Deep-sea limestones exposed in New Zealand, Italy, and Denmark show iridium increases of about 20, 30, and 160 times, respectively, above the background level at precisely the time of the Cretaceous-Tertiary extinction. Field investigations indicate that this iridium is of extraterrestrial origin, but did not come from a nearby supernova. The Alvarez hypothesis accounts for the extinction and the iridium observations. Impact of a large earth-crossing asteroid would inject about 60 times the object's mass into the atmosphere as pulverized rock; a fraction of which would remain as dust in the stratosphere for several years and be distributed worldwide. The darkness resulting would suppress photosynthesis, and the expected biological consequences match quite closely the extinctions observed in the paleontological record. One prediction of this hypothesis has been verified— the chemical composition of the boundary clay (believed to have come from the stratospheric dust) is decidedly different from that of clay mixed with the Cretaceous and Tertiary limestones, which are chemically similar to each other. The research team has made four separate estimates of the diameter of the suspected asteroid, giving a value in the range of 10 ± 4 kilometers. It has been estimated that the kinetic energy of the asteroid would have been about equivalent to that of 10^8 megatons of TNT.

In a summary of this hypothesis (Alvarez, 1980), the asteroid impact is compared with that of Krakatoa, an island volcano in the Sunda Strait between Java and Sumatra which erupted on August 26 and 27, 1883. See also **Volcano**. Whereas the estimated 14 cubic miles of material ejected into the atmosphere by Krakatoa required between 2 and $2\frac{1}{2}$ years to settle and to return the atmosphere to normal clarity, it is suggested that the debris from the hypothesized asteroid impact would have been greater by a factor of about 10^3 and thus would have put the earth essentially into darkness for a period of several years.

In 1963, when 10 Apollos were known, Öpik (Armagh Observatory in Ireland) concluded that there must be at least 43 Apollos and possibly many more. Since that time, an additional 28 Apollos have been discovered and the current rate of discovery is about four bodies per year. There is a general opinion as of the early 1980s that the number of Apollos, at a minimum, is well over 200. At this time, none of the known Apollo objects is on a collision course with the earth. However, both the Apollo and Amor asteroids are under continuous gravitational influence of nearby planets, particularly Jupiter, which causes the asteroidal orbits to precess. Because of precession, the major axis of an elliptical orbit gradually rotates through 360° in space. Thus, those asteroids with a perihelion inside the earth's orbit and an aphelion beyond the earth's orbit are destined at some time to be in an orbit that intersects the earth's orbit. It follows that the probability of any given Apollo to intersect the earth's orbit is once in every 5000 years. It further follows that the likelihood of the earth and asteroid being in precisely the same spot in the earth's orbit is very small—with an estimated collision probability of only about 5×10^9 per year (once in 200 million years).

Recent Planning for Averting Asteroid-Earth Impacts

Although not regarded too seriously by some scientists at the start, with the abatement of concerns over nuclear bomb attacks by unfriendly nations, the prevelance and risks of other catastrophic events, even from outer space, are gaining attention. The uppermost of these is asteroid-Earth collisions (impacts). From the observation of prior events and concerns, as just described in this article, many scientists now are taking seriously the program put forth by the National Aeronautics and Space Administration (NASA, U.S.) at a meeting in Los Alamos, New Mexico, in January 1992. Based upon the number of impact craters on earth, some 130 which have been identified, and the knowledge of asteroidal debris which crosses the earth's orbital path, some scientists have estimated that the average interval between impacts of kilometer-size objects is about 500,000 years. Reducing the impacting object's size down to 50 to several hundred meters, the NASA committee estimates that impacts occur at the rate of one every 200 to 300 years. The most recent occured in 1908 in Siberia, as previously described. The damage of the Siberian impact was small by comparison with the damage that could be inflicted by a 50-meter-across asteroid, releasing energy equivalent to a 15-megaton nuclear bomb. It was further postulated at the NASA meeting that a kilometer-wide impactor could release sufficient energy to annihilate a fourth of the Earth's population, depending upon location of impact and the adverse climatic and environmental effects that would follow.

Some scientists argued that a Tunguska-type (See **Mass Extinctions**) event is much more unlikely than and the effects not so great as an encounter with a swarm of sizeable asteroid particles, the latter most likely would affect very much larger Earth areas than the comparatively localized Tunguska event. Some scientists also voiced the view that swarms would be much more difficult to detect, not to mention the extreme difficulties of taking any diversionary actions.

At the Los Alamos meeting there was some serious talk pertaining to averting an asteroidal impact through careful tracking methods and possibly (1) using a nuclear bomb to break up an oncoming asteroid, or (2) using some gentler means for nudging an asteroid out of the Earth's pathway. Again, the neutron bomb would be used, exploding it off to the side of an asteroid, leaving the object intact, while radiation would heat up the asteroid surface sufficiently to vaporize it. Then the jet of vapor resulting would perform as a small rocket, thus furnishing sufficient thrust to deflect the object off a collision course. Some scientists present voiced objections to any schemes that would involve a nuclear bomb.

Possible Exploitation of Asteroids. Although in a very early stage of speculation, a number of scientists have been giving consideration to the possible use of satellites as permanent sites for space stations as well as sources of minerals. Gaffey and McCord (1977) have worked out a rather elaborate, even if preliminary plan for an asteroid mining operation, including means for transporting materials to earth. Details are given in the references listed.

Additional Reading

Beardsley, T.: "NASA Wants to Fend Off Doomsday Asteroids," *Sci. Amer.*, 30 (November 1991).

Beatty, J. K., and A. Chalken, Editors: "The New Solar System," Cambridge Univ. Press, New York, 1990.

Binzel, R. P., Gehrels, T., and M. S. Matthews, Editors: "Asteroids II," Univ. of Arizona Press, Tucson, Arizona, 1989.

Binzel, R. P., Barucci, M. A., and M. Fulchignoni: "The Origins of the Asteroids," *Sci. Amer.*, 88 (October 1991).

Bradley, J. P., and D. E. Brownlee: "An Interplanetary Dust Particle Linked Directly to Type CM Meteorites and an Asteroidal Origin," *Science*, 549–552 (February 1, 1991).

Gehrels, T., Editor: "Asteroids," Univ. of Arizona Press, Tucson, Arizona, 1979.

Kerr, R. A.: "Another Asteroid (Chiron) Has Turned Comet," *Science*, 1161 (September 2, 1988).

Kerr, R. A.: "Largest Radar Detects Dumbbell in Space," *Science*, 999 (November 24, 1989).

Kerr, R. A.: "The Great Asteroid Roast," *Science*, 527 (February 2, 1990).

Kerr, R. A.: "Impact—Geomagnetic Reversal Link Rejected," *Science*, 916 (February 23, 1990).

Kerr, A. A.: "Another Impact Extinction?" *Science*, 1280 (May 29, 1992).

Kerr, R. A.: "Did an Asteroid Leave Its Mark in Montana Bones?" *Science*, 1395 (June 5, 1992).

King, T. V. V., et al.: "Evidence for Ammonium-Bearing Minerals on Ceres," *Science*, 1551 (March 20, 1992).

Kyte, F. T., Zhou, L., and J. T. Wasson: "New Evidence on the Size and Possible Effects of a Late Pliocene Oceanic Asteroid Impact," *Science*, 241, 63–65 (1988).

MacDougal, J. D.: "Seawater Strontium Isotopes, Acid Rain, and the Cretaceous-Tertiary Boundary," *Science*, 239, 485–487 (1988).

Matthews, R.: "A Rocky Watch for Earthbound Asteroids," *Science*, 1204 (March 6, 1990).

Ostro, S. J., et al.: "Asteroid 1986 DA: Radar Evidence for a Metallic Composition," *Science*, 1399–1404 (June 7, 1991).

Powell, C. S.: "Rocky Rendezvous," *Sci. Amer.*, 20 (January 1992).

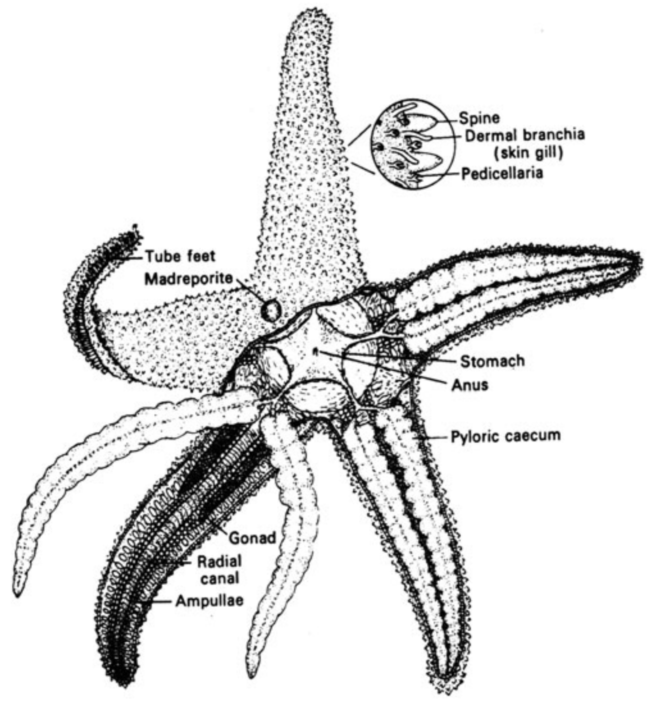
Staff: "Cuba Proposed Site for K/T Impact," *Sci. News*, 268 (April 28, 1990).

Trude, V. V., et al.: "Evidence for Ammonium-Bearing Minerals on Ceres," *Science*, 1551 (March 20, 1992).

Vilas, F., and M. J. Gaffey: "Phyllosilicate Absorption Features in Main-Belt and Outer-Belt Asteroid Reflectance Spectra," *Science*, 246, 790–792 (1989).

ASTEROIDEA. A class of the phylum *Echinodermata*. The starfishes.

The starfishes are distinguished from other echinoderms by the presence of radiating arms or rays, usually five or in multiples of five, which contain part of the internal organs and are usually not sharply separated from the central disk. There are many species but the economic importance of the group is limited. They are sometimes serious pests in oyster beds since they feed largely on shellfish.



Common starfish. (*Winchester and Lovell, "Zoology," Van Nostrand Reinhold.*)

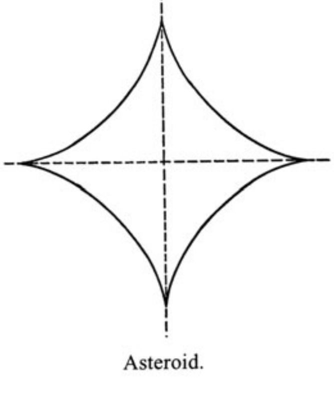
The class is divided into three orders: *Phanerozonia, Spinulosa*, and *Forcipulata*; in addition two orders—*Platyasterida* and *Hemizonida*—contain extinct asteroids.

ASTEROID (Mathematics). A higher plane curve, which is a special case of a hypocycloid. The curve is generated by a point on the circumference of a circle of radius r, which rolls around the inside of a fixed circle of radius $R = 4r$. Its parametric equations are $x = R \cos^3 \phi$, $y = R \sin^3 \phi$, and its equation in Cartesian coordinates is

$$x^{2/3} + y^{2/3} = R^{2/3}$$

The curve is symmetric to both coordinate axes. There are cusps of the first kind at the four points $(\pm R, O)$, $(O, \pm R)$, where the corresponding tangents are the X- and Y-axes. The evolute of an ellipse, which has the equation $(rx)^{2/3} + (Ry)^{2/3} = (r^2 - R^2)^{2/3}$ and the same general shape with four cusps, is sometimes also called and asteroid. The spelling astroid is often given. See accompanying diagram.

See also **Evolute** and **Hypocycloid.**



Asteroid.

ASTHENIA. Weakness, lack or loss of strength.

ASTHENOSPHERE. A term proposed by Barrell, in 1914, for the zone beneath the relatively rigid lithosphere. The asthenosphere is considered to be the level of no strain in which there is maximum plasticity, and in which the igneous rock magmas are thought to originate. See also **Earth.**

ASTIGMATIC FOCUS. In an astigmatic system (see **Astigmatism**) some of the bundle of rays from an off-axis point meet in a line perpendicular to a plane containing the point and the optical axis. Some meet in a line (at a greater image distance) which lies in a plane containing the point and the optical axis. At all other image distances the bundle is an ellipse (or circle). The first line is called the primary or meridianal or tangential focus. The second line is called the secondary or sagittal focus.

ASTIGMATISM. This term may denote: (1) A defect in a lens, including the lens of the eye, in which there is a difference in the radius of curvature of the lens as observed in one plane from that observed in another plane. (2) An aberration of a lens with spherical surfaces such that the image of a point not lying on the optical axis is a pair of short lines normal to each other and at slightly different distances from the lens. (3) In an electron-beam tube, a focus defect in which the electrons in different axial planes come to focus at different points.

ASTON WHOLE NUMBER RULE. The atomic weights of isotopes are (very nearly) whole numbers when expressed in atomic weight units, and the deviations from the whole numbers of the atomic weights of the elements are due to the presence of several isotopes with different weights.

ASTROBLEME. A scar on the surface of the earth made by the impact of a cosmic body. The term usually connotes a so-called fossil crater of ancient origin. There are 14 large and certified meteorite craters known and undoubtedly many more that are masked by vegetation or that have been subject to subsequent alteration as the result of tectonic processes, sedimentation, and erosion. Most likely the readily visible remaining craters were created during the last million years. Over the long history of the earth, some investigators believe that many thousands of giant meteorites have impacted the surface of the planet. The moon provides strong evidence that such activity in the vicinity of the earth has been strong in the past. The moon, of course, with no atmosphere and apparent minimal tectonic activity has provided a rather ideal means for permanently recording such impacts.

Particularly since the early 1980s, there has been serious and growing support for the probable impacts on earth by asteroids as well as by meteorites. Among others, the asteroid-impact hypothesis has been pioneered by Alvarez and associates (University of California, Berkeley). See **Asteroid;** and **Mass Extinctions.**

Probably the most spectacular example is the Vredefort Ring in the Transvaal of South Africa. Very little of the original crater remains, but shatter cones give evidence that this probably was the greatest terrestrial explosion in relatively recent times (within last 250 million years). At one time, geologists ascribed the structure to a series of tectonic events. A shatter cone is a distinctively striated conical fragment of rock along which fracturing has occurred, ranging in length from less than a centimeter to several meters, generally found in nested or composite groups in the rocks of cryptoexplosion structures, and generally believed to have been formed by shock waves generated by meteorite impact. Shatter cones superficially resemble cone-in-cone structure in sedimentary rocks; they are most common in fine-grained homogeneous rocks, such as carbonate rocks (limestones, dolomites), but are also known from shales, sandstones, quartzites, and granites. The striated surfaces radiate outward from the apex in horsetail fashion; the apical angle varies but is close to 90 degrees. Geologists have studied the shatter cones in the area of the Vredefort Ring and confirm that, if the rocks were returned to their original positions, the shatter cones would all point inward toward the center of the ring. It is postulated that an asteroid about a mile (1.6 kilometer) or more in diameter struck the earth from the southwest, drilling into the earth and releasing enormous shock forces. Strata some 9 miles (~14.5 kilometers) in thickness peeled back in the fashion of a flower, opening a crater some 30 miles (~48 kilometers) in diameter and 10 miles (~16 kilometers) deep. The energy released is compared with the extent of energy required to produce the Tycho and Copernicus craters on the moon. It is further estimated that the Vredefort blast was about a million times larger than the 1883 Krakatoa volcanic explosion and probably exceeded by several thousand times the largest possible earthquake. In terms of the force of nuclear explosions, it is believed that the Vredefort blast would have been classified as a 1.5-million-megaton event.

The meteorite crater (Barringer Crater) located in Arizona is much more recent (estimated 25,000 years old) and much smaller than the Vredefort event. On a nuclear scale, as mentioned in the prior paragraph, the Barringer event would have been only a 5-megaton explosion. This crater is $\frac{3}{4}$ mile (1.2 kilometers) across and 600 feet (180 meters) deep. While not in evidence at Vredefort, coesite is found at Barringer. Coesite is a monoclinic mineral, SiO_2. It is a very dense (2.93 grams/cubic centimeter) polymorph of quartz and is stable at room temperature only at pressures above 20,000 bars. The silicon has a coordination number of 4. Coesite is found naturally only in structures that are presently best explained as impact craters, or in rocks, such as suevite, associated with such structures. Coesite is believed to be a second shock-wave product and its presence has been helpful in confirming at least five astrobleme sites. Coesite was created artificially by Loring Coes, Jr. (Norton Company, Worcester, Massachusetts) in 1953 in apparatus that produced pressures exceeding 20,000 atmospheres. See also **Meteoroids** and **Meteorites.**

Coesite and suevite have been found at Ries Kessel (Giant Kettle), an ancient basin formation some 17 miles (27.4 kilometers) across and located 26 miles (41.8 kilometers) from the Steinheim Basin in southern Germany. Based upon studies within the last twenty years, Ries Kessel is now considered an astrobleme. Coesite also has been found in rather large amounts of silica glass in connection with the Wabar craters in the Empty Quarter of Arabia. Similar findings have been made at the Ashanti Crater in Ghana and at the Teapot Ess Crater in Nevada (the latter created by an atomic blast at the Nevada Proving Grounds). By seeking the presence of coesite, it is believed that additional astrobleme sites will be identified.

Several fossil craters have been identified in Canada, including Carswell Lake, Keely Lake, Deep Bay, Westhawk Lake, Lac Couture, Nastapoka Arc (Hudson Bay), Clearwater Lake, Menihek Lake, Ungava Bay, Sault-Aux-Cochons, Brent, Franktown, Lake Michikamau, Manicouagan Lake, St. Lawrence Arc (New Brunswick), Mt. Canina Crater, and Holleford. The Holleford Crater is now a slight depression about $1\frac{1}{2}$ miles (2.4 kilometers) in diameter, eroded and filled with sediments. It is located in Ontario farmland and is believed to be the result of an impact some 500 million years ago. It was discovered by means of aerial photography. Interesting shatter cone sites in the United States, in addition to the Barringer Crater, include Kentland, Indiana (in a large limestone quarry), Sierra Madera, Texas, Serpent Mound, Ohio, Flynn Creek and Wells Creek in Tennessee, and Crooked Creek in Missouri. Craters over one million years old are located at Boxhole, Dalgaranga, Henbury, and Wolf Creek in Australia.

Much pioneering work in recent years in connection with seeking shatter cones, coesite, and location of astroblemes has been done by Robert S. Dietz, whose writings on the subject are listed in the references.

Dietz suggests that the creation of coesite and of minute diamonds by meteorite impact opens up the new field of *impact metamorphism*, explaining that meteorite impacts are natural "experiments" in ultrahigh pressures on a scale that most likely will never be equaled in the laboratory.

In the twentieth century, two great impacts have been known to occur, both in Siberia. The event at Tunguska probably was caused by the fall of a comet head. At Sikhote-Alin in 1947, a very large meteorite fell that was disintegrated in mid-air, leaving more than 100 craters on the ground. All known meteorite impacts have occurred on land, but it is highly probable that many more have fallen into the sea and thus leaving evidence very difficult for geologists to uncover with present technology.

One of the most recent impact phenomena to be reevaluated is Lonar Crater, in the Buldana District of Maharashtra, India (19°58′ N, 76°31′ E). This is an almost circular depression in the basalt flows of the Deccan Traps. The crater is 1830 meters across and nearly 150 meters deep. Most of the floor is covered by a shallow saline lake (Lonar Lake). Around most of the circumference, the rim is raised about 20 meters above the surrounding plain. A second crater appears to lie about 700 meters north of the large crater. Early investigators ascribed the formation to a volcanic explosion of subsidence. However, in 1896, Gilbert emphasized the similarity of Lonar Crater with Barringer Crater in Arizona. Studies of the crater are detailed by Fredriksson, Dube, Milton, and Balasundaram in *Science*, **180**, 4088, 862–864, May 25, 1973.

ASTROGRAPHIC TELESCOPE. A refracting telescope designed to give a field of 10° or more. The objective is a designed compromise between the various optical aberrations at a specified wavelength. See also **Telescope.**

ASTROLABE. An ancient form of portable astronomical instrument invented during the second or third century B.C., probably either by Hipparchus or Apollonius. In its most common form, the astrolabe consists of a circular disk suspended by a ring so that it will hang in the plane of a vertical circle. A pointer, or alidade, is pivoted at the center of the disk, and angular graduations are marked about the edge. For purposes of measuring altitude, the ring is suspended by the thumb of one hand, and the other fingers of the same hand are employed to steady the disk as the alidade is moved, by the other hand, until it points directly at the object under observation. The altitude can then be read directly on the disk.

The astrolabe was used by navigators for the determination of latitude from the fifteenth century until the invention of the sextant. Since that time, it has been used as a teaching instrument in elementary classes. The astrolabe, in its modern version, is essentially the only impersonal instrument for the measurement of time and latitude that does not rely on secondary standards. The zenith telescope will do the same observational tasks as the astrolabe, but it is necessary to introduce nonfundamental stars. The modern astrolabe is free of personal errors, and gives stellar positions with an accuracy on the order of one-tenth of a second of arc.

ASTROMETRY. The branch of astronomy dealing with the positions, distances, and motions of the planets and stars; it includes determination of time and position. See also **Bonner Durchmusterung.**

ASTRONAUTICS. The science of travel in outer space.

Weight and Weightlessness. In Newtonian mechanics, weight is understood to mean the force that an object exerts upon its support. This would depend on two factors: the strength of gravity at the object's location (things weigh less on the moon) and, as Newton called it, the quantity of matter in a body (its "mass"). At any given location, where gravity is fixed, mass can be measured relative to a standard by noting the extension of a spring to which it and the standard are successively attached. Alternatively, the unknown and standard may be hung at opposite ends of a rod and the balance point noted. However, by an entirely separate experiment, mass can also be measured by noting the resistance of the object to a fixed force applied horizontally on a frictionless table. The measured acceleration provides the required basis of comparison with the standard. Needless to say, all objects measure identical accelerations when freely falling in the vertical force of gravity. This merely means that, unlike the arbitrary force we apply horizontally in the experiment above, gravity has the property of adjusting itself in just the right amount, raising or lowering its applied force, to maintain the acceleration constant.

It was well known that objects appear to increase or decrease their weight (alter the extension of the spring) if the reference frame in which the measurement takes place accelerates up or down. As gravity did not really change, however, most people were inclined to draw a distinction between *weight* defined as mg, where m is the mass and $\mathbf{g}$ is the local gravity field, and the *appearance of weight*, the force of an object on its support as measured by the spring's extension. One way to avoid the difficulty has been to speak of an *effective* $\mathbf{g}$, which takes into consideration the frame's acceleration. For example, at the equator of the earth, we measure, say by timing the oscillation of a pendulum, the effective $\mathbf{g}$, some 0.34% less than the $\mathbf{g}$ produced by the mass of earth beneath our feet. If the earth were rotating with a period of an hour and a half instead of 24 hours, our centripetal acceleration at the equator would cause the effective $\mathbf{g}$ to vanish completely, our scales would not register, objects would be unsupported, and for all practical purposes we would be weightless.

Formally, we could state that any accelerating frame produces a local gravitational field $\mathbf{g}_{\text{acc}}$ that is equal and opposite to the acceleration. Thus, a rotating frame generates a centrifugal $\mathbf{g}_{\text{acc}}$ opposing the centripetal acceleration. We have at any point

$$\mathbf{g}_{\text{eff}} = \mathbf{g} + \mathbf{g}_{\text{acc}} \tag{1}$$

where $\mathbf{g}$ is the field produced by matter along (e.g., the earth). By identical reasoning, an object in orbit, whether falling freely in a curved or in a straight path, will carry a reference frame in which $\mathbf{g}_{\text{eff}}$ is zero, for its acceleration will always exactly equal the local $\mathbf{g}$ by the definition of the phrase, "freely falling."

This concept was placed on a firm footing by Einstein who maintained that Eq. (1) is reasonable not only in mechanics but in all areas of physics including electromagnetic phenomena. We arrive at the inevitable conclusion that we cannot distinguish by any physical experiment between an apparent $\mathbf{g}$ accountable to an accelerating frame and a "real" $\mathbf{g}$ derived from a local accumulation of mass. This central postulate of the General Theory of Relativity also unified the two separate conceptions of mass. An object resting on a platform that is accelerating toward it will resist the acceleration in an amount depending on its inertia. It presses against the platform with a force equal to that it would have if placed at rest on the surface of a planet with local field equal and opposite to the acceleration of the frame.

General Principles of Central Force Motion. The gravitational force between point masses is inverse square, written

$$mg = -\frac{\gamma m' m}{r^2} \tag{2}$$

where the center of coordinates from which the unit vector $\hat{\mathbf{r}}$ is described lies in m', one of the masses. Thus, the force on m is directed $-\hat{\mathbf{r}}$, toward m' and is proportional to $1/r^2$ with γ the constant of proportionality. The quantity $\mathbf{g}$ is the force on *m divided by m*, (or normalized force) for which the name "gravitational field of *m′*" is reserved. Of course, if m were in the field of a collection of mass points, or even in a continuous distribution of mass, the summated or integrated $\mathbf{g}$ at the location of m would no longer be an inverse square function with respect to any coordinate center. However, in one special case, the inverse square functional form would be preserved: if the source mass were symmetrically distributed about the coordinate center. This would be the case if the source were a spherical shell or solid sphere, of density constant or a function only of r. The sun and earth can be regarded, at least to a first approximation, as sources of inverse square gravitational fields.

There are some important general statements we can make about the motion of an object placed with arbitrary position and velocity in a centrally directed force field, i.e., a field such as the one described, which depends only on distance from a central point (regardless of whether or not the dependence is inverse square). As the force has only a radial and no angular components, it cannot exert a torque about an axis through the center. This means that the initial angular momentum is conserved. Now angular momentum is a vector quantity and therefore is conserved both in direction and magnitude. It is defined by $\mathbf{r} \times \mathbf{p}$, where $\mathbf{r}$ is the position vector to the mass of momentum $\mathbf{p}$. The direction of the angular momentum vector is thus perpendicular to the plane containing $\mathbf{r}$ and $\mathbf{p}$. As this direction is permanent, so also must be the plane. The planar motion of the object can be expressed in polar coordinates, so that, by writing $\mathbf{r} = r\,\hat{\mathbf{r}}$ and $\mathbf{p} = m(r\,\hat{\mathbf{r}} + r\,\dot{\phi}\hat{\boldsymbol{\phi}})$, we find the specific angular momentum (angular momentum per unit mass) called h, to be

$$h = r^2\dot{\phi} \tag{3}$$

This too then must be a constant of the motion.

Consider now the rate at which area is swept out by the radius vector, dS/dt. We recall from analytic geometry that $dS = \frac{1}{2}r^2\, d\phi$. Thus

$$\frac{dS}{dt} = \frac{h}{2} \tag{4}$$

so that this is a constant of the motion as well. On integration, we conclude that the size of a sector that is swept out is proportional to the time required to sweep it out. In the case of a closed orbit, the total area S would then be related to the specific angular momentum as

$$S = \frac{hT}{2} \tag{5}$$

This sector area-time relationship is Kepler's second law of planetary motion which was induced from Tycho Brahe's observation of Mars without prior knowledge of gravity and its central character.

The Laws of Kepler. Kepler stated two other laws of planetary motion: The orbits of all the planets about the sun are ellipses (a radical departure from the circles of Copernicus), and the squares of their period are proportional to the cubes of their mean distance from the sun, this mean being the semimajor axis of their ellipses. See Fig. 1. The third law pertained to the one object common to all the planets, the sun. Taken together, the three laws led Newton to the concept of gravitational force and its inverse-square form.

By applying Newton's law of motion $\mathbf{F} = m\mathbf{a}$, a relationship between $\mathbf{a}$, the second derivative of the position vector, expressed in polar form, and $\mathbf{F}/m$ or $\mathbf{g}$, as given by Eq. (2), leads to the familiar conic solution for the trajectory of an object in an inverse square field,

$$\frac{1}{r} = \frac{\gamma m'}{h^2} + A\cos(\phi - \phi_0) \tag{6}$$

where A and ϕ_o are constants. A rotation of axis will eliminate ϕ_o, thereby aligning the coordinate axis with the conic's major axis. Also, by expressing the general conic, an ellipse or hyperbola, in terms of the usual parameters of semimajor axis a and eccentricity $\in$, we can relate the geometric parameters to the gravitational-dynamical constants, viz:

$$h = [\gamma m' a(1 - \epsilon^2)]^{1/2} \tag{7}$$

and

$$\frac{1}{r} = \frac{\gamma m'}{h^2}(1 + \epsilon\cos\phi) \tag{8}$$

Note that by substituting Eq. (7) into Eq. (5) and expressing the area of an ellipse as $S = \pi a^2(^1 - \epsilon^2)^{1/2}$ we arrive at Kepler's third law,

$$T = \frac{2\pi}{(\gamma m')^{1/2}}\, a^{3/2} \tag{9}$$

The *energy* of the orbiting object can be calculated with ease by evaluating it at an extremal point, say the nearest point to the gravitational source, called pericenter or perifocus. As the energy is constant, it is immaterial where the calculation is made. Here the velocity has only an angular component so that the kinetic energy for a unit orbiting mass is $\frac{1}{2}v^2 = \frac{1}{2}r^2\dot{\phi}^2$. The potential energy at pericenter is $-\gamma m'/r_{\text{pe}}$ where r_{pe} is the distance of the unit mass from m', the focal point. Here $\gamma = 0$ so that by Eq. (8),

$$\frac{1}{r_{\text{pe}}} = \frac{\gamma m'}{h^2}(1 + \epsilon) \tag{10}$$

On substituting Eq. (7), we find the total kinetic and potential energy to be

$$E = \frac{\gamma m'}{2a} \tag{11}$$

Our conclusion: All objects in orbit with the same major axes have identical periods and identical energies per unit mass. Knowledge of E is invaluable in determining an object's speed when its distance from the source is known and vice versa.

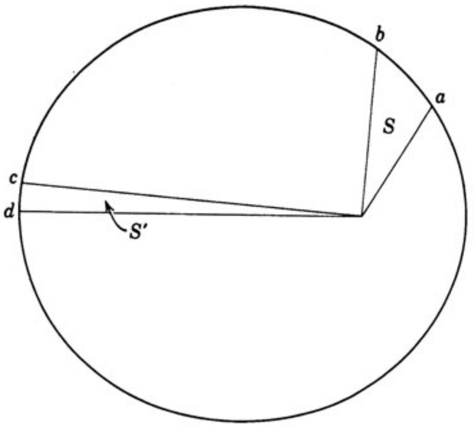


Fig. 1. Kepler's second law. The sector area S swept out is proportional to the time required for the planet to move from a to b. Thus, if $t_{cd} = t_{ab}$, then $S' = S$.

In the event that the orbiting object's mass is not negligibly small compared with that of the gravitational source, one must take note that the combined center of mass, from which the acceleration is described, no longer may be assumed to lie in the center of the gravitational source. This complicates our equations somewhat, for the accelerating force still is expressed relative to the center of the source (if spherical). The adjustment that results, when center of mass coordinates are transformed to relative coordinates in the expression for acceleration, requires our equations to take the form $\gamma(m' + m)$ wherever formerly $\gamma m'$ appeared. See Fig. 2.

Disturbances in the Central Field. The Earth, of course, is spherical only to a first approximation. More accurately, it is an ellipsoid of revolution about a minor axis—an oblate spheroid. Still more accurately, it appears to be slightly pear-shaped and, in addition, its figure is distorted by continuous local variations. The spheroidal figure, nevertheless, accounts for nearly all the anomalous effects of satellite orbits. For one thing, the gravitational force on the satellite is no longer centrally directed; the excessive mass in the equatorial plane produces

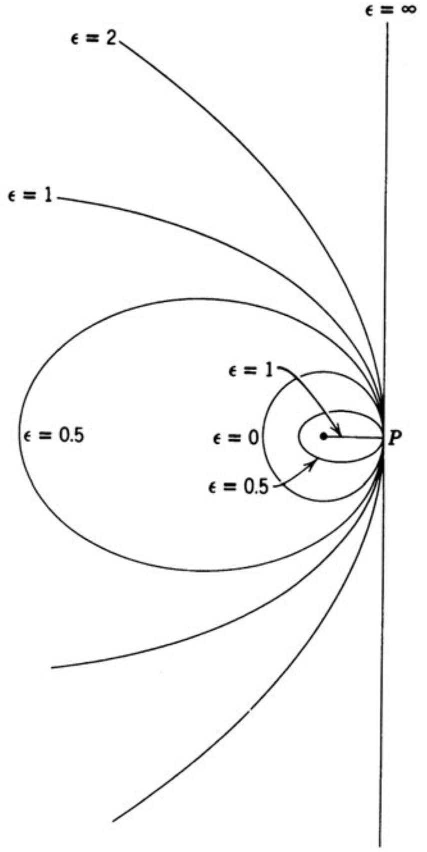


Fig. 2. Orbits of differing eccentricities and major axes which pass through a common point. Higher speeds correspond to higher energies and longer major axes.

a force on the satellite directed out of its orbital plane. The resultant torque causes the direction of the angular momentum vector to change, i.e., the plane containing the satellite's ellipse turns. The plane turns continuously about the polar axis maintaining its angle with the axis and with the equatorial plane constant. The turning rate is greatest for low orbits and small angles of inclination with the equator. For polar satellites, the plane remains fixed. A separate effect of this equatorial bulge perturbative force is the slow turning of the ellipse's major axis *within* the orbital plane. This effect vanishes at an inclination of 63.4°; the major axis turns backward at inclinations above this angle and forward below. See Fig. 3.

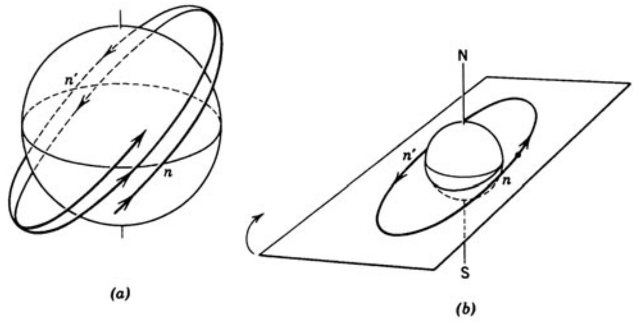


Fig. 3. The orbit of an earth satellite. The earth's equatorial bulge causes retrograde motion of the points of intersection n and n' of the orbit and equatorial plane. This can alternatively be interpreted as a retrograde motion, about the north-south axis, of the plane containing the closed orbit. The plane moves in the direction shown by the arrow in (b), maintaining a constant angle with the axis.

Rocket Propulsion. A rocket operates by the simple principle that if a small part of its total mass is ejected at high speed, the remaining mass will receive an impulse driving it in the opposite direction at a moderate speed. As δm_e, the propellant, leaves at speed v_e with respect to the rocket, the remaining rocket mass m receives a boost in speed δv such that

$$\delta m_e v_e = m \delta v \tag{12}$$

If additional equal propellant mass is ejected at the same speed, the boost in rocket speed is slightly greater than before as the rocket mass has been slightly depleted by the prior ejection. Indeed, if the residual rocket mass eventually were minuscule, its boost in speed could reach an enormous value. The integrated effect of these nonlinear boosts is found as

$$v_t - v_o = v_e \log_e \frac{m_0}{m_t} \tag{13}$$

where v_0 and m_0 are the rocket speed and mass at some arbitrary initial time and v_t and m_t are the same quantities at some time t later.

From these simple considerations, it is apparent that the highest rocket velocities are attained if we could increase the propellant speed as well as the mass ratio m_0/m_t. The mass ratio can be maximized by obvious methods such as choosing a high-density propellant which cuts the tankage requirement or avoiding unnecessarily complicated apparatus for ejecting propellant at high speed. A nuclear rocket, for example, may perform well in its ability to eject propellant an order of magnitude higher in velocity than conventional chemical rockets; nevertheless, the penalty required in reactor weight and shielding severely limits its effectiveness.

Specific impulse is one performance characteristic which applies to the propellant's ability to be ejected at high speed regardless of the weight penalty required to do this. It is the impulse produced per mass of propellant ejected, or $m\ \delta v/\delta m_e$, or, by Eq. (12), simply v_e. In engineering usage, it is impulse per *weight* of propellant ejected, or v_e/g_e where g_e is the acceleration of gravity at the earth's surface. Its units are seconds, and it can be interpreted as the thrust produced by a rocket per weight of propellant ejected per second. By itself, thrust is of little importance unless it is sustained for a significant time by a large backup of propellant tankage. It is here that the mass ratio term in Eq. (13) would play an important role in any evaluation of a rocket's true performance.

Transfer Orbits. If one wishes to leave one orbit and enter another by rocket, an optimum path is generally chosen to minimize the total propellant required. Nevertheless, this should not be done at the expense of unduly long flight times, complicated guidance equipment, or high acceleration stresses. These would require unprofitable weight expenditures which would offset the frugality in propellant tankage. See Fig. 4.

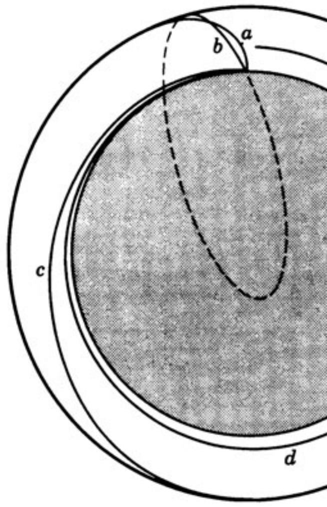


Fig. 4. Four launch trajectories into a satellite orbit about a planet. (a) If that planet has an atmosphere, the rocket may ascend in a "synergic" trajectory from the planetary surface to the final orbit, i.e., it cuts through the denser portions in an initially vertical path and gradually bends over into a horizontal path during burnout. (b) If there is no atmosphere it may ascend from the ground in a ballistic ellipse. This same ascent path may be chosen if the departure is from a parking orbit or "space platform" close to ground level. A far better choice would be (c) the Hohmann ellipse, with pericenter at the planet's surface and apocenter at the satellite orbit. Burnout time is assumed short in both this and the ballistic case. (d) A vehicle such as an ion rocket, which can sustain a microthrust for a very long time, cannot be launched from the ground but only from a parking orbit. It will spiral out to the desired altitude with few or many turns about the planet, depending on the magnitude of the thrust relative to that of the gravitational force.

Let us examine a simple but recurring example of a transfer problem, that of leaving a space platform in one circular orbit and entering another larger one concentric with the first. If the transfer path were radial or near radial (a so-called ballistic orbit) then one would have to launch at a large angle to the direction of motion of the platform, accomplished only by a velocity component opposed to the platform's motion. On reaching the outer platform, a soft landing can be made only by a substantial rocket velocity boost tangent to the orbit. Clearly, the total propellant expenditure would be far greater than one alternative of launching the rocket in the direction of motion of the first platform with just sufficient speed to reach the outer circle, timed so that the outer platform will meet the spacecraft. The transfer orbit will be an ellipse cotangent with both circles. The outer platform will be moving much faster of course at the contact point as the major axis of its orbit is much greater [see Eq. (11)], but the difference in speed is not nearly as pronounced as for the ballistic transfer case. A differential speed increment at contact completes the maneuver.

The return trip, from an outer to inner circle, is made by following the second half of this cotangent ellipse, named the Hohmann transfer orbit after the German engineer who discovered its optimal property with regard to propellant expenditure. In the return case, the spacecraft is launched in opposition to the outer platform's motion. This removes kinetic energy and forces the spacecraft to fall in closer to the attractive center in order to make cotangent contact with the inner circle. The total propellant expenditure from the outer to the inner platform is the same as for the original journey.

An interesting question arises if one wishes to leave a platform for an outer orbit when it initially is in an elliptical orbit rather than a circle. Should we depart from apocenter where we are furthest from the gravitational source and closest to our destination? Or should we depart instead from some other point in the ellipse? Paradoxically, our best launch point is at pericenter, for here the largest possible amount of energy will be transferred to the spacecraft for a given expenditure of propellant. A given thrust applied for a given time interval will do more work on the spacecraft when it is moving fast, as at pericenter, for it covers a greater distance during the interval. This advantage offsets the undesirability of being at a lower potential energy point at pericenter.

Powered Trajectories. In the usual operation of a solid- or liquid-propelled rocket, the propellant is depleted in a time negligibly small compared with the total flight time. The trajectory analysis may generally be considered as that of a free orbit subject to burnout initial conditions as in the discussion above. If, however, the propellant ejection is sustained over long periods, as in an ion-propelled rocket, the trajectory analysis is necessarily complicated, for, in addition to the varying gravitational force, the vehicle, of slowly diminishing mass, is subject to a thrust which may be changing both in direction and magnitude. Even one of the simplest thrust programs, a constant thrust in the direction of motion, requires an electronic computer analysis in order to obtain the position and velocity at future times.

The continuous-thrust trajectory is a spiral with many advantages over the orbital ellipses. First, the lower sustained thrust precludes the high acceleration stresses associated with rapid-burning chemical rockets. Much of the structural weight usually needed to withstand these stresses can be replaced by propellant. Also, flights to the extremities of a gravitational region may take a shorter time in a spiral trajectory. In a long Hohmann ellipse, for example, most of the journey is made at very low speed. In a powered spiral, on the other hand, the spacecraft could be made to move fast, for the thrust, though small, is integrated over many months.

The spiral concept is ideal for rockets where very high ejection velocities are feasible by using electromagnetic or electrostatic particle accelerators, but only at the expense of a low propellant flow rate and relatively heavy power-generating equipment. However, the propellant reserve, and thrust, could then last the required long time. Such an ion rocket with its very low thrust-to-weight ratio could hardly be expected to take off from the ground, and could only take off from an orbital platform. In the vacuum of space, the ion beam meets its ideal environment.

ASTRONOMICAL CLOCK. A clock that indicates astronomical events as well as time. Historically, these clocks were developed to achieve reliable timekeeping by the mechanical simulation of the observable astronomical relationships of celestial bodies. Probably the first such clock was the great Chinese astronomical clock tower of Su Sung (1020–1101) which incorporated the first solarsideral gear and used (on the water wheel that drove the clockworks) an escapement mechanism believed to be the world's first. The first recorded astronomical clock in Europe was produced about 1330 by Richard of Wallingford of the Abbey of St. Albans in Hertfordshire, England. Better known as the "Astrarium," completed in 1364 by Giovanni de Dondi of Chioggia, Italy. An exact reproduction of this device is in the Smithsonian Institution's Museum of History and Technology, Washington, D.C. This clock indicated the movements of the Sun, Moon and the five then-known planets and displayed mean solar time and a perpetual Julian calendar for the Church's movable feasts. The outstanding modern example of an astronomical clock is Jens Olsen's "World Clock" in Copenhagen's City Hall. Detailed visual simulation of complex astronomical events within and beyond the solar system is now far more accurately achieved by planetarium projectors. See **Planetarium.**

ASTRONOMICAL UNIT. A unit of distance principally employed in expressing distances within the solar system, but also used to some extent for measuring interstellar distances. Technically defined, one astronomical unit is the mean distance of the earth from the sun. To express this in linear units, it becomes necessary to determine the distance of the earth from the sun in the units chosen or, in other words, to determine the solar parallax. The value for the length of the astronomical unit is 1.495985×10^8 kilometers (92.956×10^6 miles), and was obtained by radar astronomy. See also **Light Year;** and **Parsec.**

ASTRONOMY. In broad terms, modern astronomy may be defined as the science of matter and energy in outer space and more particularly concerned with the composition, mass, relative position, size, and other chemical and physical properties of celestial bodies, including asteroids, clusters, comets, galaxies, meteroids, natural satellites, nebulae, planets, stars, intervening space and dust, and all other forms of cosmic material and phenomena. Astronomy, like many of the other traditional mother sciences has, in recent years, been segmented into a number of specialties, such as astronometry, astrophysics, cosmogony, cosmology, among others.

In this encyclopedia, there are nearly 400 separate articles that relate to some specific topic in astronomy, including astronomical catalogs and directories; constellations; coordinate systems, motions, paths, positions, forces, and units; meteorites and comets; planets and planetoids, stars and galaxies; telescopes and other astronomical instrumentation systems; and space-exploration programs principally directed toward obtaining astronomical knowledge. See accompanying table. Additional scores of articles throughout the encyclopedia contain information that relates in some way to astronomy. The majority of the aforementioned separate articles incorporate lists of references for additional reading. For maximum utility of this encyclopedia, frequent reference to the comprehensive alphabetical index should be made.

New Astronomy—Progress Report

In the last (1989) edition of this encyclopedia, the term "New Astronomy" was introduced to describe a number of fundamental changes and approaches employed by the astronomy community, in contrast with traditional practices that had identified the science over the past several decades (from the 1960s to the mid-1980s). These were summarized as follows:

> Traditionally, astronomy has been an observational rather than an experimental science, i.e., observing cosmic effects and then theoretically deducing their causes. Increasingly precise instrumentation has immensely enhanced what is "seen" and in recording what has been seen, not simply on a photographic plate, but also in digitized displays. For many decades, the visible portion of the electromagnetic spectrum was the astronomer's principal probe into space, gaining entrance into space by way of the optical telescope. During the last few decades, nearly all other bands of the electromagnetic spectrum (x-ray, gamma-ray, infrared, ultraviolet, radio frequency, etc.) have been utilized and are playing an increasingly important role. It should be stressed, however, that the optical telescope remains a very useful tool. Knowledge of the solar system, of course, has advanced from largely qualitative descriptions to quantitative information as gleaned by the various earth-launched space probes to several planets— and satellites have been used to explore far beyond the solar system by mounting sensitive instruments on so-called earth-orbiting space platforms. As dramatic as the foregoing developments appear to be, a somewhat less glamorous advancement has been occurring in astronomical research, namely, the use of earth-bound and what could be called laboratory-based experiments. Mathematical modeling, coupled with computer analysis and simulation, today is enabling the astronomer to "compress or expand" time so to speak and to explore cosmological theories and arguments, wherein the observational effect-cause relationship can become a computer devised cause-effect relationship—then testing the latter against the real effects in space that are observed. Against the background of centuries-old astronomical observations, progress made through vastly improved instrumentation and the application of computer technology is but in its infancy as of the late 1980s. See Fig. 1.

Some of the aspirations of and plans of astronomers expressed in the late 1980s have been achieved or partially achieved by early 1994; other projects have been put on hold; still others have brought disappointment. The program to incorporate new optical and imaging techniques and to construct entirely new facilities for astronomical-optical telescopes generally has proceeded according to schedule, notably the Keck instruments on Mauna Kea, Hawaii. Some other projects are somewhat behind schedule. The status of important optical installations are reported in some detail in article on **Telescope (Astronomical-Optical),** including the Very Large Telescope (VLT) at the European Southern Observatory (ESO), the National New Technology Telescope (NNTT), the Columbus Telescope (Mount Graham, Arizona), and others.

The space orbiting telescope, Hubble, has been disabled since its launch in 1990 and is scheduled for repair in space toward the end of

ASTRONOMICAL TOPICS DESCRIBED IN THIS VOLUME

Astronomy and Related Disciplines

- Astronometry
- Astronomy
- Astrophysics
- Cosmogony
- Cosmology

Catalogs and Directories

- Almagest
- Almanac (Astronomical)
- Bonner Durchmusterung
- Star Finder

Constellations

- Andromeda
- Aquarius (the water bearer)
- Aquila (the eagle)
- Aries (the ram)
- Auriga (the charioteer)
- Bootes (the herdsman)
- Camelopardalis (the giraffe)
- Cancer (the crab)
- Canis Major (the great dog)
- Canis Minor (the little dog)
- Capricornus (the sea-goat)
- Carina
- Cassiopeia (the chair)
- Centaurus (the centaur)
- Cepheus
- Cetus (the whale)
- Chamaeleon
- Columbia (the dove)
- Constellations
- Corona Australis (southern crown)
- Corona Borealis (northern crown)
- Corvus (the crow)
- Crater (Constellation)
- Crux (southern cross)
- Cygnus (the swan)
- Draco (the dragon)
- Eridanus
- Gemini (the twins)
- Grus (the crane)
- Hercules
- Horologium
- Hydra (the serpent)
- Indus
- Leo (the lion)
- Lepus (the hare)
- Libra (the scales)
- Lupus (the wolf)
- Lyra (the harp)
- Monoceros
- Ophiuchus
- Orion (the hunter)
- Pegasus (the winged horse)
- Perseus
- Phoenix
- Pisces (the fishes)
- Piscis Austrinus
- Puppis
- Reticulum (Constellation)
- Sagitta (the arrow)
- Sagittarius (the archer)
- Scorpius (the scorpion)
- Sextans
- Taurus (the bull)
- Telescopium
- Triangulum (the triangle)
- Triangulum Australe
- Tucana
- Ursa Major (the greater bear)
- Ursa Minor (the smaller bear)
- Vela
- Virgo (the virgin)

Coordinate Systems, Motions, Paths, Positions, Forces, Units

- Antisolar Point
- Aphelion
- Apparent Horizon (See **Horizon (Celestial)**)
- Astronomical Triangle
- Astronomical Unit
- Autumnal Equinox (See **Equinox**)
- Azimuth (Astronomy)
- Barycentric Parallax
- Bode's Relation
- Celestial Mechanics
- Celestial Sphere
- Conjunction (Astronomy)
- Declination
- Distance Modulus
- Doppler Effects
- Eccentricity Correction
- Eclipse
- Ecliptic
- Einstein Shift
- Elongation (Astronomy) (See **Conjunction (Astronomy)**)
- Equatorial Coordinates (Astronomy)
- Equinox
- Evection
- Galactic Coordinates
- Geocentric Coordinates (Astronomy)
- Geocentric Parallax
- Horizon (Astronomical)
- Horizon (Celestial)
- Horizontal Coordinate System (Astronomy)
- Hour Angle
- Inclination
- Inertial System (The Primary)
- Kepler's Laws of Planetary Motions
- Latitude
- Librations
- Light-Year
- Line of Apsides
- Line of Nodes
- Mach Principle
- Meridian
- Nadir
- Nutation
- Occultation
- Opposition (Astronomy) (See **Conjunction (Astronomy)**)
- Orbit (Astronomy)
- Parallax (Astronomy)
- Parsec
- Perihelion
- Perturbation (Astronomy)
- Position Angle (Stellar)
- Precession (Astronomy)
- Quadrature (Astronomy)
- Relativity and Relativity Theory
- Retrograde Motion
- Secular Parallax
- Secular Terms
- Solar Parallax
- Solstice
- Sphere of Influence (Planet)
- Syzygy

Coordinate Systems, Motions, Paths, Positions, Forces, Units

- Three-Body Problem (Astronomy)
- Transit (Astronomy)
- Two-Body Problem (Astronomy)
- Vertical Circle
- Zenith
- Zodiac

Meteorites and Comets

- Ablation (Meteorite)
- Achondrites
- Aerolite
- Astrobleme
- Bennett (Comet)
- Bielids
- Bolide
- Chondrite
- Comet
- Earthpoint
- Encke (Comet)
- Fireball
- Halley's Comet
- Kohoutek (Comet)
- Leonids
- Lyrids
- Meteor
- Meteorite
- Meteorite Crater
- Meteoroid
- Meteor Shower
- Meteor Trail
- Micrometeorite
- Perseids
- Poynting-Robertson Effect
- Radiant Point (Meteor)
- Shooting Star
- Tektite

Planets and Planetoids

- Albedo (Astronomy)
- Asteroid
- Ceres (See **Asteroid**)
- Davida (**See Asteroid**)
- Earth
- Earthlight
- Eros
- Eunomia (See **Asteroid**)
- Gegenschein
- Heat Balance (Planet)
- Hektor (See **Asteroid**)
- Jovian
- Jupiter
- Mars
- Mercury (Planet)
- Moon (The)
- Neptune
- Pallas (See **Asteroid**)
- Planets and the Solar System
- Pluto
- Satellite (Planetary)
- Saturn
- Uranus
- Venus
- Vesta (See **Asteroid**)

Space-Exploration Programs

- Apollo Program (See **Moon**)
- Pioneer Missions to Planets
- Space Astronomy
- Space Station and Space Shuttle
- Viking Mission to Mars
- Voyager Mission to Jupiter and Saturn

Stars and Galaxies

Specific Stars and Galaxies

- Achernar (Alpha Eridani)
- Adhara (Epsilon Canis Majoris)
- Aldebaran (Alpha Tauri)
- Algol (Beta Persei)
- Alpha Centauri
- Alpha Crucis
- Alpheratz (Alpha Andromedae)
- Altair (Alpha Aquilae)
- Antares (Alpha Scorpii)
- Arcturus (Alpha Bootes)
- Bellatrix (Gamma Orionis)
- Beta Centauri
- Beta Crucis
- Betelgeuse (Alpha Orionis)
- Canopus (Alpha Carinae)
- Capella (Alpha Aurigae)
- Castor (Alpha Geminorum)
- Fomalhaut (Gamma Piscis Austrini)
- Hyades
- Milky Way
- Mira

(Continued)

ASTRONOMICAL TOPICS DESCRIBED IN THIS VOLUME *(Continued)*

Mizar (Zeta Ursae Majoris)
Pleiades
Polaris (Alpha Ursae Minoris)
Pole Star (See **Vega (Alpha Lyrae)**)
Pollux (Beta Geminorum)
Praesepe (See **Cancer (The Crab)**)
Procyon (Alpha Canis Minoris)
Regulus (Alpha Leonis)
Rigel (Beta Orionis)
Shaula (Lambda Scorpi)
Sirius (Alpha Canis Majoris)
Spica (Alpha Virginis)
Sun
Vega (Alpha Lyrae)
Virgo Cluster

Types of Stars and Galaxies
Binary Stars
Black Hole
Cepheids
Cluster (Galactic)
Double Star
Eclipsing Binary
Flare Stars
Galaxies
Giant and Dwarf Stars
Magellanic Clouds
Navigators' Stars
Nebula
Neutron Stars
Nova and Supernova
Planetary Nebula (See **Nebula; Lyra (The Harp)**)
Quasars
Radio Pulsars
Ring Nebula (See **Nebula**; **Lyra (The Harp)**)
Seyfert Galaxy (See **Galaxy**)
Spectral Class (See **Star**)
Spectroscopic Binaries
Spiral Galaxy (See **Andromeda; Galaxy**)
Star
Supernova 1987A
Universe (See **Cosmology**)
Variable Star
Visual Binaries

Stellar and Galactic Properties and Phenomena
Bailey Beads
Color Index
Color Temperature (Stellar)
Effective Temperature (Astrophysics)
Einstein Shift
Green Flash
Insolation
Interstellar Molecules
Interstellar Reddening
Limb Darkening
Mass-Luminosity Relation
Period Luminosity Law
Proper Motion (Star)
Purple Light
Radial Velocity (Star)
Red Shift
Scintillation (Astronomical)
Space Velocity (Stellar)
Spectral Equivalent Width (Stellar)
Stellar Associations

Stars and Galaxies
Stellar Curve of Growth
Stellar Heat Index
Stellar Luminosity
Stellar Magnitude
Stellar Parallax
Sunshine (See **Solar Energy**)
Talcott Pairs (See **Zenith Telescope**)
Twilight
Velocity Curve (Stellar)
Zodiacal Light

Telescopes and Other Astronomical Instrumentation Systems

Alidade (See **Astrolabe**)
Altazimuth
Angular Measurement (Eccentricity Correction)
Artificial Horizon
Astrographic Telescope
Astrolabe
Augmentation
Bolometer (See **Stellar Heat Index**)
Catadioptic System (See **Maksutov-Bouwers Telescope**)
Coronagraph (See **Sun**)
Coudé
Dynamical Parallax
Eccentricity Correction
Filar Micrometer
Galilean Telescope
Gamma-Ray Astronomy
Gauss Eyepiece
Gregorian Telescope
Heliostat
Infrared Astronomy
Maksutov-Bouwers Telescope
Meridian Circle
Light Pollution
Newtonian Telescope
Objective Prism
Photometry (See **Stellar Magnitude**)
Radio and Radar Astronomy
Ramsden Circle
Refraction (Astronomical)
Resolving Power (Telescope)
Ritchey-Chretien Telescope
Schmidt Objective
Schwarzschild Telescope
Seeing (Astronomy)
Semidiameter Correction (Sextant)
Spectroheliograph (See **Sun**)
Spectroscopic Parallax
Telecentric System
Telescope (Optical-Astronomical)
Ultraviolet Astronomy
X-Ray Astronomy
Zenith Telescope

Time-Related Topics
Apparent Solar Time (See **Time**)
Apparent Sidereal Time (See **Time**)
Astronomical Clock
Atomic Time (See **Time**)
Calendar
Coelostat
Ephemeris Time (ET) (See **Time**)
Equation of Time
Heliostat (See **Sun**)
Mean Solar Time (See **Time**)
Saros
Sidereal Period
Synodic Period
Time
Universal Time (UT) (See **Time**)

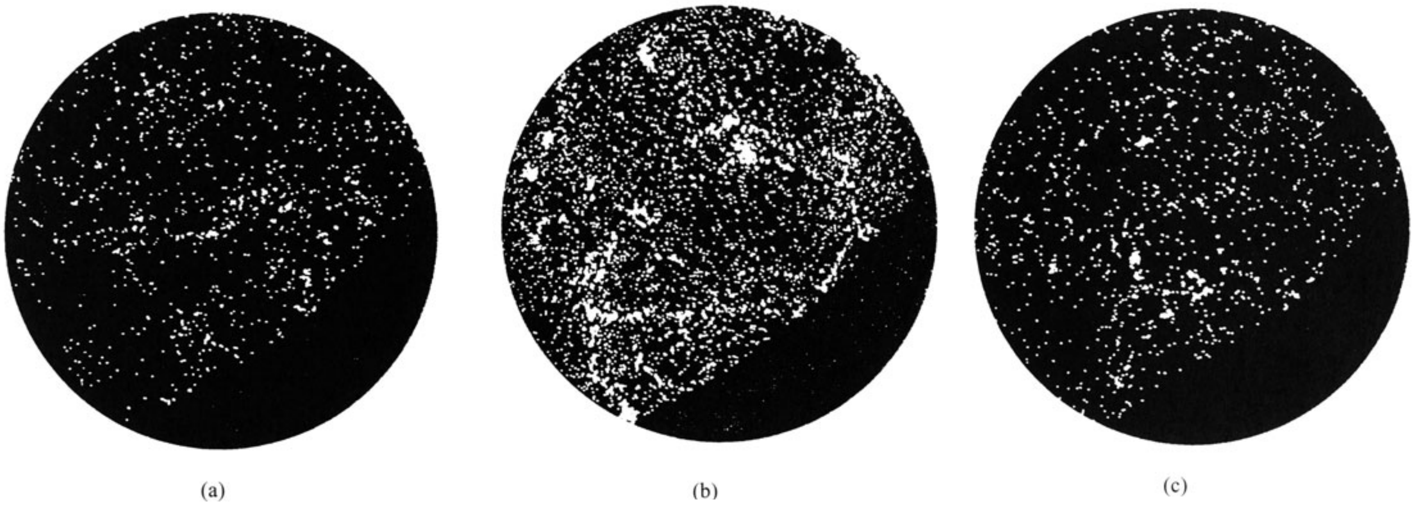


Fig. 1. Photofacsimiles of computerized simulations of large-scale structure in the universe. Points represent galaxies. (a) Clustering of galaxies on large scale as observed; (b) a plausible initial condition if universe were filled with sufficient numbers of light neutrinos to account for all dark matter—not an exceptionally good match with (a); (c) a universe dominated by cold, dark matter as it was formed appears to provide a better match with (a). Simulations of this type have been carried out by scientists from the University of California, Cambridge, and the University of Arizona, among others. (*Ref. Krauss, 1986.*)

Fig. 2. Picture of Gaspra asteroid taken by the Galileo Spacecraft from a range of 5300 kilometers (3300 mi), about 10 minutes before closest approach on October 29, 1991.

Surprisingly, Gaspra's surface is apparently covered with granular material similar to soil, called regolith. Variations in brightness and color on the surface show that some of the granular material may have moved from higher ridges to lower depressions. The presence of a regolith is unexpected on such a small body because it should be very easy for the particles to escape Gaspra's gravitational pull. More study of this surface layer using measurements still stored on the spacecraft will give important information about Gaspra's composition.

Another important result, the rate of crater production in the inner Solar System, comes from determining the number of craters on the surface of Gaspra. Compared with most planetary satellites, Gaspra shows a low crater density, especially for medium and large craters (2–6 kilometers, or about 1–4 miles, in diameter). The number of craters increases sharply for smaller-sized craters. Belton and his collaborators conclude from this information that collisions by small objects are more common in the inner Solar System than had been believed. The conclusion can also be applied to the origin of small crater production on the Earth's moon. One explanation of the numbers of small lunar craters has been that many are caused by secondary impacts as debris from larger collisions falls back to the surface. It now seems that the Moon's surface, like Gaspra's, may have been subject to more small-object collisions than previously thought.

Gaspra's age can also be deduced from the number of craters on its surface. The Galileo scientists estimate that Gaspra has existed as an asteroid for about 200 million years. An asteroid of Gaspra's size should survive without major collisions for about 500 million years on average. It is thought that Gaspra was formed during a collision of larger bodies.

Estimates of Gaspra's age and cratering rate are dependent on the material from which it is made. If Gaspra is made up mostly of metal, it is likely to be much older, since it will be more stable against collisions. The calculations were made assuming that the asteroid has a stony composition. Asteroids similar to Gaspra are thought to be the parent bodies of stony meteorites, the most common type of meteorite found on earth. Another possibility is that Gaspra and similar asteroids have large amounts of metal, making them parents of the rarer stony-iron meteorites.

Gaspra has a mean diameter of 14 kilometers (8.7 miles), with length of about 16 kilometers (10 mi) and a width of 12 kilometers (7.4 mi). It rotates with a period of 7.04 hours. Gaspra's diameter places it in size between the two satellites of Mars—Deimos, with a diameter of 6.2 kilometers (3.8 mi), and Phobos, with a diameter of 11.1 kilometers (6.9 mi)—both of which are used as comparison bodies in studying Gaspra. (*National Optical Astronomy Observatories and NASA photo.*)

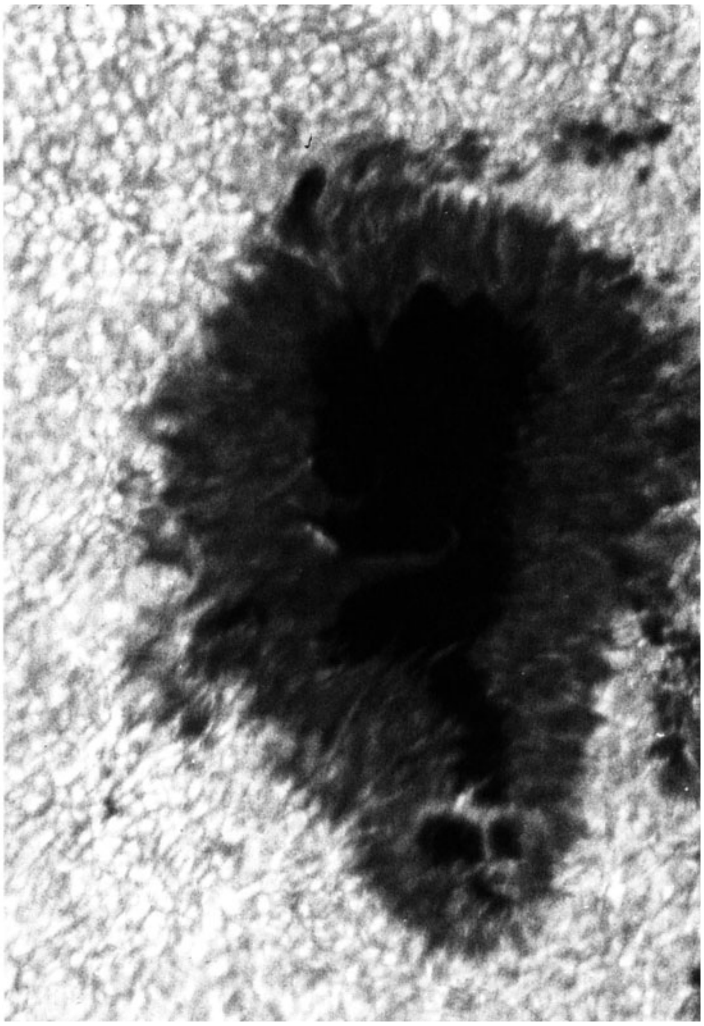


Fig. 3. Photo of recently formed sunspot taken at the McMath-Pierce solar facility on Kitt Peak, near Tucson, Arizona.

The unusually large sunspot appeared after a period of little sunspot activity. This image was taken on October 1, 1993. Later photos show the sunspot to have separated into two distinct spots. The sunspot is about 40,000 miles across, equivalent to about five earth diameters. This photo was taken by solar astronomers William Livingston and Jack Harvey of the National Solar Observatory.

The 11-year sunspot cycle reached a maximum in 1992 and is expected to be at a minimum in 1996.

Several unique solar incidents have been photographed by Dr. Livingston over the course of his 34-year career as a Kitt Peak astronomer, using the McMath telescope, which began operating in 1962. The McMath-Pierce is responsible for more active research than any other solar facility in the world, and the three-mirror Arizona telescope continues to provide excellent solar data. The telescope is also used at night for stellar and planetary research programs. (*Photo: National Optical Astronomy Observatories.*)

1993. The results of that effort are reported in the aforementioned article.

Much attention during the 1990s has been given to the concepts of adaptive and active optics, also described in the aforementioned article.

Probably the most devastating disappointment to astronomers was the loss of signals from the *Mars Observer,* which turned silent on August 21, 1993, just short of its goal. See article on **Mars.** In addition to reporting in specific articles, two recently reported projects are dramatically shown in Figures 2 and 3. Some hint of what may be expected from the astronomy sciences during the next half-decade may be found in the list of papers scheduled for the immediately forthcoming (February 1994) of the American Association for the Advancement of Science:

1. Cosmology after COBE (Cosmic Background Explorer) satellite.
2. Changing Perspectives on the Planets—Results from the Magellan Mission.
3. Cosmic Rain—The Bombardment of Earth.
4. Gravitational Biology and Space Medicine—Results from experiments in space and ground simulations.

ASTROPHYSICS. Commencing with the advent of photography and the study of stellar spectra in the second half of the nineteenth century, astrophysics now includes optical and radio observations of stars, clusters, interstellar material, galaxies and clusters of galaxies, and their interpretations. Radiation from these external sources provides information on the direction of the source, its velocity, composition, temperature, and other physical conditions, including magnetic fields, density, degree of ionization, and turbulence. The term *astrophysics* is generally understood to include all these aspects except the measurement of direction (positions of stars in the sky and changes due to parallax and proper motion), and the orbits of planets, asteroids and comets (celestial mechanics). Because of its proximity, the sun can be studied in more detail than other stars (solar physics); its structure and its influence on the nearby planets are closely related to geophysics and stellar astrophysics. Study of the motions of stars in pairs, groups, clusters, associations, and galaxies is the overlap of celestial mechanics with astrophysics, and the study of the distribution and patterns of motion of the distant galaxies is the overlap with cosmology.

ASYMMETRIC TOP. A model of a molecule which has no threefold or higher-fold axis of symmetry, so that during rotation all three principal moments of inertia are in general different. Example, the water molecule.

ASYMMETRY (Chemical). Asymmetry involves the presence of four different atoms or substituent groups bonded to an atom. Its existence was discovered in 1815 by the French physicist, J. B. Biot (1774–1867). Biot found that oil of turpentine and solutions of sugar, camphor, and tartaric acid all rotate the plane of plane-polarized light when placed between two Nicol prisms. This phenomenon is called *optical rotation* and is indicated in symbols, such as: $[\alpha]_D^{20^\circ} = +53.4$ aq., signifying that the substance gives a rotation of 53.4° to the right (clockwise, or plus) in water solution at 20°C using sodium D line as the light source. Substances in solution that rotate light to the right are designated *d* and are called *dextrorotatory*; substances rotating light to the left are designated *l* and are called *levorotatory*. See also **Isomerism.**

ASYMMETRY. See **Conservation Laws and Symmetry.**

ASYMPTOTE. The limiting position of a tangent to a curve, where the point of contact is only at an infinite distance from the origin. Where there are no infinite branches, as in the cases of the circle and the ellipse, there is no real asymptote.

Suppose the equation of a given curve can be expanded in a power series

$$y = f(x) = \sum_{k=0}^{n} a_k x^k + \sum_{k=1}^{\infty} b_k / x^k = S_1 + S_2$$

Then, if $\lim_{x \to \infty} S_2 = 0$, the equation of the asymptote is $y = S_1$. If this equation is linear, the asymptote is a straight line; otherwise, it is a more complicated curve. In the linear case, the equation of the asymptote may be written as

$$y = mx + b; \quad m = \lim_{n \to \infty} f'(x)$$

$$b = \lim_{x \to \infty} [f(x) - xf'(x)]$$

ASYMPTOTIC RELATIVE EFFICIENCY (or ARE). The efficiency of an estimator of a statistical parameter (as compared with an optimal estimator) as the sample size on which the estimator is based tends to infinity.

ASYMPTOTIC SERIES. A divergent series of the form

$$A_0 + A_1/x + A_2/x^2 + \cdots + A_n/x^n + \cdots$$

It is an asymptotic representation of a function *f(x)* if

$$\lim_{x \to \infty} x^n [f(x) - S_n(x)] = 0$$

for any value of n, where S_n is the sum of the first $(n + 1)$ terms of the series.

A familiar example of an asymptotic series is the Euler-Maclaurin formula, which converges for a certain number of terms and then begins to diverge. If one includes a large number of terms in this formula the successive derivatives become increasingly larger in the numerator and they increase much more rapidly than the coefficients, which occur in the denominator. However, if the summation is stopped with the term just before the smallest and not with the smallest term, the error is usually about twice the neglected term. Thus one can obtain satisfactory results in this case and with other such series when they are used with caution. Other examples are the logarithmic integral and the gamma function, both of which can be developed as asymptotic series.

An asymptotic expansion is unique; that is, a given function can be represented by only one such series. It may be integrated, two or more of them can be multiplied together, but in general it should not be differentiated.

ASYNCHRONOUS. This is a term used to designate the property of a device or action whose timing is not a direct function of the clock cycles in the system. In an asynchronous situation, the time of occurrence or duration of an event or operation is unpredictable due to factors such as variable signal propagation delay or a stimulus which is not under control of the computer. See also **Synchronous.**

In terms of a computer channel, an asynchronous channel does not depend upon the computer clock pulses to control the transmission of information to and from the input or output device. Transmission of the information is under the control of interlocked control signals. Thus, when a device has data to send to the channel, the device activates a service request signal. Responding to this signal, the channel activates a "service out" signal. The latter, in turn, activates a "service in" signal in the device and also deactivates the request signal. Information then is transferred to the channel in coincidence with "service in" and the channel acknowledges receipt of the data by deactivating "service out."

Asynchronous operation also occurs in the operation of analog-to-digital subsystems. The system may issue a command to the subsystem to read an analog point and then proceed to the next sequential operation. The analog subsystem carried out the A/D conversion. When the conversion is complete, the subsystem interrupts the system to signal the completion.

Asynchronous also has a broader meaning—specifically unexpected or unpredictable occurrences with respect to a program's instructions.

ATACAMITE. This mineral is a basic chloride of copper corresponding to formula $Cu_2Cl(OH)_3$. Crystallizes in thin, orthorhombic prisms, may occur massive. Hardness, 3–3.5; sp gr, 3.76–3.78; luster, adamantine to vitreous; color, green, streak, green; transparent to translucent.

It is a secondary mineral found associated with malachite and cuprite; originally found at Atacama, Chile, whence its name. Other localities are Bohemia, South Australia, and in the United States in Arizona, Utah, and Wyoming. See also **Cuprite;** and **Malachite.**

ATAVISM. The appearance through heredity of characters which have not been developed in the parents of the organism in question. The strict meaning of the word is the reappearance of grandparental characters, but it has been used also to designate the reappearance of characters from more remote generations.

ATAXIA. Lack of muscular coordination due to disease of the brain and nervous system, particularly the cerebellum or spinal cord. Occurs in cerebral palsy. Degeneration of portions of the spinal cord in later uncontrolled stages of syphilis (*neurosyphilis*) will cause loss of coordination of the limbs (*locomotor ataxia*), much less frequently seen where there is an active public health program directed to the detection, treatment, and prevention of venereal diseases.

Hereditary ataxias may develop in disorders with a known metabolic basis, but the majority of inherited cerebellar and spino-cerebellar degenerations are of unknown causation. Those of early onset are usually of autosomal recessive inheritance, e.g., Frederich's ataxia, while later-onset cases of cerebellar degeneration are most often dominantly inherited.

R. C. V.

ATAXIC. A term applied by Keyes, in 1901, to all unstratified ore deposits in contradistinction to sedimentary, stratified or eutaxic ore deposits.

ATELECTASIS. Collapse of part, or the whole, of a lung. This may be congenital, as in the stillborn infant whose lungs have never been expanded by the act of breathing; more commonly it is acquired, resulting from obstruction to a bronchus by a mucous plug, especially after surgical operations; occasionally by pressure from without as from bony deformity or tumor growth. Atelectasis is a prominent feature of adult respiratory distress syndrome (ARDS).

ATHERMAL TRANSFORMATION. A reaction that occurs without thermal activation. Such a reaction also takes place without diffusion and can occur with great rapidity under the influence of a sufficiently high driving force. The martensite transformation that occurs in steel is primarily athermal, so that the amount of austenite transformed to martensite depends primarily on the temperature to which the steel is cooled and not upon the rate of cooling or the length of time the metal is held at the quenching temperature. It is necessary to note the difference between an isothermal transformation and an athermal transformation. In the former, the reaction occurs at constant temperature and depends, in general, on both diffusion and thermal activation. The transformation of austenite to pearlite can occur isothermally, with carbon atoms diffusing out of the austenite and into the cementite lamellae. See also **Iron Metals, Alloys, and Steels.**

ATHEROSCLEROSIS. See **Arteries and Veins; Ischemic Heart Disease.**

ATLANTIC SUITE. A term proposed by A. Harker, in 1896, for the chemically and structurally related igneous rocks of the Atlantic coast line. Chemically the rocks of this suite are described as alkaline and are represented by such types as granite and its magmatic relatives, as compared with the calc-alkali igneous rocks of the Pacific Suite.

ATMOLYSIS. The separation of a mixture of gases by means of their relative diffusibility through a porous partition, as burned clay. The rates of diffusion are inversely proportional to the square roots of the densities of the gases. Hydrogen, thus, is the most diffusible gas.

ATMOSPHERE (Earth). An envelope (actually a series of envelopes) in the form of imperfect spherical shells of various materials that are bound to the earth by gravitational force. Consisting of gases, vapors, and suspended matter, the total mass of the earth's atmosphere is estimated at approximately 5.1×10^{15} tons, or somewhat less than one-millionth part of the total mass of the earth. One-half of this total mass lies below about 5500 meters (18,000 feet). More than three-fourths of the atmosphere exists below about 10,700 meters (~35,000 feet). The composition of the lower layers of the atmosphere is assumed for purposes of most engineering calculations as 76.8% nitrogen and 23.2% oxygen by weight; 79.1% nitrogen and 20.9% oxygen by volume. A more precise composition of this mixture of gases, including minor constituents, is given in entry on **Air.**

The earth's atmosphere extends some 600 to 1500 kilometers into space. Two factors are involved in this great extension of the atmosphere. First, above about 100 kilometers, the atmospheric temperature increases rapidly with altitude, causing an outward expansion of the atmosphere far beyond that which would occur were the temperature within the bounds observed at the earth's surface. Second above this distance, the atmosphere is sufficiently rarefied so that the different atmospheric constituents attain diffusive equilibrium distributions in the gravitational field; the lighter constituents then predominate at the higher altitudes and extend farther into space than would an atmosphere of more massive particles. This effect is enhanced by the dissociation of some molecular species into atoms.

The Challenge of Atmospheric Research

The truly scientific study of the earth's atmosphere is a relatively recent phenomenon as compared with most of the fundamental sciences. Although philosopher-scientists like Benjamin Franklin and Thomas Jefferson shed light on weather processes in the 18th Century, and while the invention of the telegraph made possible the first accurate mapping of weather patterns, the theoretical study of the atmosphere did not begin until the late 19th Century, and the rapid development of atmospheric research did not begin until after World War II (late 1940s). Since then, theoretical advances have occurred at a steadily increasing rate, supported by high-speed computers; by vehicles such as instrumented aircraft, high-altitude balloons, rockets, and satellites; and by new sensors, such as radars, lasers, and instruments for measuring the many chemicals present in the atmosphere today.

Atmospheric science presently is recognized as containing some of the most difficult and challenging problems that confront any science. The atmosphere is now viewed as a very complex mixture of chemical and physical processes, linked with the atmosphere of the sun, with the oceans and the earth's plant and animal life, and, of course, with the continually growing effects of human activities (deforestation, pollution, etc.).

A major advancement in atmospheric science in the United States was the establishment, in 1960, of the National Center for Atmospheric Research (Boulder, Colorado), NCAR, which, in turn, has impacted atmospheric science worldwide. The formation of NCAR dates back to 1956 when the Committee on Meteorology (later renamed the Committee on Atmospheric Sciences) of the National Academy of Sciences recommended an increase of 50 to 100% in support for basic meteorological research and the establishment of NCAR to be operated by a consortium of universities with federal support from the National Science Foundation. As of the late 1980s, scientists and technologists who specialize in atmospheric science from 50 U.S. universities, including the Universities of Alaska and Hawaii; 2 Canadian universities (McGill and the University of Toronto); and 3 other institutions (the Naval Postgraduate School, Scripps Institution of Oceanography at the University of California, San Diego, and the Woods Hole Oceanographic Institution), constitute the staff of NCAR. The operating entity of NCAR is the University Corporation for Atmospheric Research (UCAR).

In commenting on the university consortium and inter-institution concept, Roscoe Braham, who heads the cloud physics research program at the University of Chicago and on the UCAR Board of Trustees stated in 1985, "I think it was clear from the outset that the problems and opportunities in meteorology were so enormous that individual university departments could not cope with their magnitude. In the late 1950s, many of us were dissatisfied with the slow progress that was being made in weather forecasting. Weather modification was viewed as a major opportunity area. The *Thunderstorm Project* (a joint research effort by several U.S. federal agencies based at the University of Chicago) in 1947 had shown how valuable aircraft and radar could be in studying the atmosphere, but the Air Force, which supplied these tools and other equipment for that project, could no longer provide that kind of large-scale support for basic research in the universities. Big computers existed, but nobody had them, practically speaking. There was a general feeling that if we—the universities—had access to resources bigger than ourselves, we could do a much better job of basic research on the atmosphere."

An abridged, but representative list of NCAR targets would include:

1. *Storms*—analyzing and predicting mesoscale weather. An outstanding weather problem is the unanticipated, sometimes destructive, always hazardous occurrence of small-scale weather disturbances. This is exemplified by the demise of Delta Air Lines Flight 191 at Dallas-Fort Worth International Airport on August 2, 1985 when the aircraft encountered a microburst (downdraft). See also article in this encyclopedia on **Fronts and Storms.**
2. *Atmospheric Chemistry*—gaining new insights on the composition of the atmosphere. When one reviews the knowledge of atmospheric chemical composition as it existed two or three decades ago, one is struck by the primitive state of the science at that time. The atmosphere near the earth was viewed as a fluid in motion, transporting moisture and heat. It also transported pollutants arising from cities, factories, and fires. The chemical species in the air were regarded as essentially inert and for good reason—most of the components that were known were inert gases. It is

now understood that the atmosphere is a reactive environment. See also **Pollution (Air).**

3. *Relationship of Sun with Earth*—gaining greater knowledge of the sun, not so much from an astronomical viewpoint, but rather how the earth reacts to changes on the sun. A great deal of research has focused on the solar corona and sunspots. By way of helioseismology, measurements of motions in the interior of the sun have provided a better understanding of the solar activity cycle. As pointed out by Robert Noyes (Harvard University), "This is one of the most important issues in solar physics, especially in terms of solar-terrestrial relationships. Almost every effect of the sun on the earth is magnetically induced through the sunspot cycle, whether it's the influence of solar flares on the upper atmosphere and the aurora or the effect of fluctuations in the ultraviolet flux from the sun on ozone formation in the upper atmosphere."
4. *Observing and Modeling the Global Atmosphere*—the atmosphere is the prototypical chaotic nonlinear system. This was shown by the simplest atmospheric model, devised by Edward Lorenz in the mid 1960s, the starting point for modern mathematical studies of such systems. Because the atmosphere is chaotic, atmospheric models are sensitive to small variations in initial conditions and possess an inherent growth of error. These properties impose a theoretical limit on the range of deterministic predictions of large- scale flow patterns of about two weeks. As early as 1735, these complications were recognized by George Hadley, an English lawyer and spare-time scientist, who stated, "I think the causes of the General Trade-Winds have not been fully explained by any of those who have wrote on that subject." See also article on **Climate.**
5. *Application of Modern Computer and Instrument Technology*—the first computer was acquired by NCAR in 1964 (CDC 3500) and, over the years, has greatly enhanced its computer capabilities, acquiring a CRAY-1, in 1977 and a second CRAY-1 in 1983. A CRAY X-MP/48 is the latest acquisition and will be used for more realistic climate simulations, thunderstorm and tornado modeling, three-dimensional chemical-dynamical models for studying problems, such as acid precipitation, and models of the solar cycle and the general circulation of the ocean. NCAR also has a generous complement of airborne equipment, radar, and balloons, the oldest yet still valuable tool for reaching up into the atmosphere, measuring it, and bringing samples back or recording them. Huge scientific balloons in current use can carry payloads (in thousands of pounds) to altitudes of 100,000 feet (30 km) and higher. See also article on **Balloon.**

Other subjects in this encyclopedia related to the aforementioned topics are listed at the end of this article.

Atmosphere-Altitude (Pressure-Temperature) Relationships

The composition of the atmosphere does not change much up to 100 kilometers; there is a region of maximum concentration of ozone (still a very minor constituent) near 20 to 30 kilometers; the relative concentration of water vapor falls markedly from its average sea-level value up to 10 or 15 kilometers, and the relative abundance of atomic oxygen begins to become appreciable on approaching 100 kilometers, due to photodissociation of oxygen by ultraviolet sunlight. Above 200 kilometers, atomic oxygen is the principal atmospheric constituent for several hundred kilometers. However, helium is even lighter than atomic oxygen, so its concentration falls less rapidly with altitude, and it finally replaces atomic oxygen as the principal atmospheric constituent above some altitude which varies with the sunspot cycle between 600 and 1500 kilometers. At still higher altitudes, atomic hydrogen finally displaces helium as the principal constituent. The hydrogen extends many earth radii out into space and constitutes the telluric hydrogen corona, or *geocorona*.

The temperature of the upper atmosphere, and hence its density, varies with the intensity of solar ultraviolet radiation and this, in turn, varies with the sunspot cycle and with solar activity in general. The solar radio-noise flux is a convenient index of solar activity, since it can be monitored at the earth's surface. The minimum nighttime temperature of the upper atmosphere above 300 kilometers has been expressed in terms of the 27-day average of the solar radio-noise flux at 8-centimeter wavelength. This varies from about 600 K near the minimum of the sunspot cycle to about 1400 K near the maximum of the cycle. The maximum daytime temperature is about one-third larger than the nighttime minimum.

Various properties of the earth's atmosphere are described in Tables 1 through 5 and by Figs. 1 and 2. The several layers of the atmosphere are indicated in Table 1, along with the relationship between atmospheric pressure and altitude. Atmospheric density versus altitude are given in Table 2. Geopotential altitude as related with actual altitude and the acceleration due to gravity is given in Table 3. It is interesting to note that the energy required to lift an object 2 million geometric feet is only 1.824 million times that required to lift it 1 foot above sea level— this because of the decrease in the acceleration due to gravity with altitude.

Reduction of molecular weight, indicating the change in composition of the atmosphere with increasing altitude, is shown in Table 4. The molecular weight of air is assumed essentially constant from sea level up to about 300,000 feet (91,440 meters). At altitudes higher than this, lower molecular weight is largely attributed to the dissociation of oxygen. Above an altitude of about 590,000 feet (179,832 meters), the lower molecular weight is also affected by the diffusive separation and dissociation of nitrogen.

The percent water vapor content of air at saturation versus representative temperatures and pressure altitudes is given in Table 5.

The layers of the earth's atmosphere of interest to meteorologists are the *troposphere* and the *stratosphere*. The troposphere is a thermal atmospheric region, extending from the earth's surface to the stratosphere and characterized by decreasing temperature with height, appreciable vertical wind motion, appreciable water vapor content, and containing nearly all clouds, storms, and pollutants. The thickness of the troposphere varies from as little as about 7–8 kilometers in the cold polar regions to more than 13 kilometers in the warmer, equatorial regions. Temperatures decrease to the interface between the troposphere and stratosphere. This interface is termed the *tropopause*. At the tropopause, polar temperatures average around −55°C, in equatorial regions, −80°C. Above the stratosphere are the *mesosphere* and *ionosphere*, and the outermost layer, the exosphere, gradually fades into the plasma continuum between earth and sun.

In these higher layers of the atmosphere, complex interactions between the fluxes of electromagnetic radiation of various wavelengths and corpuscular radiation from the sun on one side and the low-density concentrations of atmosphere gases on the other side take place. The particulate radiations are also governed by the earth's magnetic field. Radiations of short wavelength cause a variety of photochemical reactions, the most notable of which is the creation of a layer of ozone acting as an effective absorber of solar ultraviolet and thus causing a warm layer at 30 kilometers in the atmosphere. See **Aerosol;** and **Oxygen.** The upper atmosphere, as an absorber of primary cosmic rays, shows many interesting nuclear reactions and is an important natural source of radioactive substances, including tritium and carbon 14 which are used as tracers of atmospheric motions and as criteria of age. See also **Climate.**

Most manifestations of weather take place in the troposphere. They are governed by the general atmospheric circulation which is stimulated by the differential heating between tropical and polar zones. The resulting motions in the air are subject to the laws of fluid dynamics on a rotating sphere with friction. They are characterized by turbulence of varying time and space scale. Evaporation of water (see Table 5) from the ocean and its transformation through the vapor state to droplets and ice crystals, forming clouds and precipitation, are important symptoms of the weather-producing forces.

The term *ecosphere* is sometimes used to identify that part of the lower atmosphere where unaided breathing is possible. In meteorology, the term upper atmosphere is sometimes used. That part of the atmosphere above the lower troposphere is called the *upper air*, for which no distinct lower limit is set, but the term is generally applied to the levels above 850 millibars.

The ionosphere is described in a separate entry, **Ionosphere.**

Heat Balance in the Atmosphere

Total heat received directly from the sun, at the outer limits of the atmosphere (the amount that would be received at the earth's surface if

TABLE 1. ATMOSPHERIC PRESSURE VERSUS ALTITUDE ABOVE SEA LEVEL

	Altitude Above Sea Level		Pressure	
Atmospheric Layer	feet (thousands)	meters (thousands)	inches of mercury	millibars
	2000	609.60	7.959×10^{-12}	269.524×10^{-12}
Mesosphere G	1920	585.22	10^{-11}	3.4×10^{-10}
	1320	402.34	10^{-10}	3.4×10^{-9}
	1000	304.80	5.256×10^{-10}	177.989×10^{-10}
	900	274.32	10^{-9}	3.4×10^{-8}
F	640	195.07	10^{-8}	3.4×10^{-7}
Ionosphere	480	146.30	10^{-7}	3.4×10^{-6}
	400	121.92	10^{-6}	3.4×10^{-5}
	340	103.63	10^{-5}	3.4×10^{-4}
E	300	91.44	10^{-4}	3.4×10^{-3}
	260	79.25	10^{-3}	3.4×10^{-2}
D	200	60.96	10^{-2}	3.4×10^{-1}
Chemosphere	140	42.67	10^{-1}	3.4
	100	30.48	0.32	10.8
	95	28.96	0.4	13.5
	90	27.43	0.5	17.1
	85	25.91	0.64	21.6
	80	24.38	0.81	27.4
Ozonosphere	75	22.86	1.03	34.9
	70	21.34	1.31	44.4
	65	19.81	1.67	56.6
	60	18.29	2.12	71.8
	55	16.76	2.69	91.1
	50	15.24	3.42	115.8
Stratosphere	45	13.72	4.35	147.3
	40	12.19	5.54	187.6
	35	10.67	7.04	238.4
	30	9.14	8.89	301.1
	25	7.62	11.10	375.9
	20	6.10	13.75	465.6
Troposphere	15	4.57	16.89	572.0
	10	3.05	10.58	696.9
	5	1.52	24.90	843.2
	0	0	29.92	1013.3

Conversion factors used: 1 foot = 0.3048 meter

passage were unaffected by the atmosphere and clouds) is very nearly 1.94 gram-calories per square centimeter per minute. This great quantity of heat is distributed in such a way that the maximum is received directly below the sun, with a decreasing amount received as the distance from the heat equator increases. It is for this reason, of course, that tropical areas are warm; polar regions are cold.

Not all the sun's radiation is received at the earth's surface. Clouds and snow reflect about 75% of solar radiation falling upon them; land surfaces reflect an average of 10–30%; water reflects varying percentages, from 70% when the sun is only 5° high, to less than 2% when the sun is over 50° above the horizon. Some solar radiation is absorbed by the atmosphere gases and some by water vapor in the air. Another part is lost to the earth by scattering in the atmosphere. Altogether, solar radiation is distributed as follows: (1) approximately 42% is sent back into space by reflection; (2) 15% is absorbed by the atmosphere and its impurities and cloud particles; and (3) 43% is received and absorbed by the earth's surface. On cloudy days (average cloudiness is about 52%), considerably less solar radiation reaches the earth than on clear days. Loss on a clear day is approximately 17% of the total amount; but on a clouded day, the loss is about 78%. Deserts are conspicuously clear, and therefore receive a much larger percentage of the incoming solar heat than do continental west coasts, which have considerable cloud cover. Snow-covered regions lose a larger percentage of their incoming solar heat than do forest- and vegetation-covered lands. Water surfaces, averaged the world over, do not reflect a large percentage of solar heat, but water is capable of absorbing large quantities of heat with only a small temperature change. The influence of local terrain on solar radiation plays a considerable role in determining the daily and seasonal temperatures of that area.

The earth receives its heat from a number of sources: (1) about 17% is direct solar radiation; (2) 10% is sky radiation (from scattered solar radiation); (3) 70% is long-wave radiation received from the atmosphere surrounding the earth; and (4) 3% is received by contact with warm surface air currents. It should be realized, however, that all this energy, regardless of its immediate source, originates from the sun.

The fact that there is no accumulation of heat on the earth indicates a radiative heat balance. Radiation received by the earth is dispersed as follows: (1) 7% goes to space by radiation through transparent bands in the atmosphere (transparent to radiation from a black body at 300°K); (2) 78% goes to the atmosphere by radiation, where it is absorbed and redistributed; and (3) 15% is used in evaporation processes and is carried to the atmosphere, where it adds to the store of atmospheric heat. Water vapor is the principal absorber of earth radiation as it passes through the atmosphere. Carbon dioxide and ozone also have some strong absorption bands.

Those regions between approximately 35°N and 35°S receive more energy than they radiate back to space, whereas the other regions of the earth receive less energy than they radiate. The excess of energy from the subtropical and tropical zones is transferred toward the poles by both the ocean currents and the atmospheric winds. The advection of heat energy balances the differential in direct radiation. Thus, the average temperature at any point on the earth remains sensibly the same from year to year.

TABLE 2. ATMOSPHERIC DENSITY VERSUS ALTITUDE ABOVE SEA LEVEL

Altitude Above Sea Level		Specific Weight	
feet (thousands)	meters (thousands)	pounds per cubic foot	kilograms per cubic meter
2000	609.60	1.614×10^{-15}	25.856×10^{-13}
1000	304.80	2.374×10^{-13}	38.031×10^{-13}
100	30.48	0.00101	0.016
95	28.96	0.00129	0.021
90	27.43	0.00166	0.027
85	25.91	0.00214	0.034
80	24.38	0.00275	0.044
75	22.86	0.00350	0.056
70	21.34	0.00445	0.071
65	19.81	0.00566	0.091
60	18.29	0.00720	0.115
55	16.76	0.00915	0.146
50	15.24	0.01164	0.186
45	13.72	0.01480	0.237
40	12.19	0.01883	0.302
35	10.67	0.02370	0.380
30	9.14	0.02861	0.458
25	7.62	0.03427	0.549
20	6.10	0.04075	0.653
15	4.57	0.04812	0.771
10	3.05	0.05648	0.905
5	1.52	0.06590	1.056
0	0	0.07648	1.225

Conversion factors used: 1 foot = 0.3048 meter
1 pound/cubic foot = 16.02 kilograms/cubic meter.

Elasser's radiation chart is one of the better known charts for the graphical solution of the radiative transfer problems important in meteorology. Given a radiosonde record of the vertical variation of temperature and water vapor content, such quantities can be found with this chart as the effective terrestrial radiation, net flux of infrared radiation at a cloud base or a cloud top, and radiative cooling rates. A chart of this type used widely in Europe is the Möller chart.

Atmospheric radiation is infrared radiation emitted by or being propagated through the atmosphere. Atmospheric radiation, lying almost entirely within the wavelength interval of from 3 to 80 micrometers, provides one of the most important mechanisms by which the heat balance of the earth-atmosphere system is maintained. Infrared radiation emitted by the earth's surface is partially absorbed by the water vapor of the atmosphere, which, in turn, re-emits it, partly upward, partly downward. This secondarily emitted radiation is then, in general,

TABLE 4. MOLECULAR WEIGHT OF ATMOSPHERE VERSUS ALTITUDE

Altitude Above Sea Level		
feet (thousands)	meters (thousands)	Molecular Weight
2000	609.6	15.67
1900	579.12	15.80
1800	548.64	15.96
1700	518.16	16.13
1600	487.68	16.33
1500	457.20	16.56
1400	426.72	16.82
1300	396.24	17.14
1200	365.76	17.51
1100	335.28	17.97
1000	304.8	18.54
900	274.32	19.27
800	243.8	20.24
700	213.36	21.59
600	182.88	23.60
500	152.4	24.09
400	121.92	24.76
300	91.44	28.89
0	0	28.97

Conversion factor used: 1 foot = 0.3048 meter.

TABLE 5. PERCENT WATER VAPOR CONTENT OF AIR AT SATURATION VERSUS REPRESENTATIVE TEMPERATURES AND PRESSURE ALTITUDES

Temperature °C	1,000 Millibars 370 Feet (113 Meters)	850 Millibars 4,780 Feet (1,457 Meters)	700 Millibars 9,880 Feet (3,011 Meters)	500 Millibars 18,280 Feet (5,572 Meters)
40	4.97%	5.93%	7.35%	—
30	2.76	3.28	4.03	5.79
20	1.49	1.76	2.16	3.06
10	0.77	0.91	1.12	1.57
0	0.38	0.45	0.55	0.77
−10	0.18	0.21	0.26	0.36
−20	0.08	0.09	0.11	0.16
−30	0.03	0.04	0.05	0.06
−40	0.01	0.01	0.02	0.02

TABLE 3. ACCELERATION DUE TO GRAVITY AND GEOPOTENTIAL ALTITUDE VERSUS ACTUAL ALTITUDE ABOVE SEA LEVEL

Actual Altitude Above Sea Level		Geopotential Altitude		Acceleration Due to Gravity	
feet (thousands)	meters (thousands)	feet (thousands)	meters (thousands)	feet/second/ second	meters/second/ second
2000	609.6	1825	556.26	26.79	8.17
1800	548.64	1657	505.05	27.26	8.31
1600	487.68	1485	452.63	27.75	8.46
1400	426.72	1310	399.29	28.25	8.61
1200	365.76	1132	345.03	28.77	8.77
1000	304.8	950	289.56	29.3	8.93
800	243.8	766	233.48	29.84	9.1
600	182.88	579	176.48	30.4	9.27
400	121.92	389	118.57	30.97	9.44
200	60.96	196	59.74	31.57	9.62
0	0	0	0	32.17	9.81

Conversion factors used: 1 foot = 0.3048 meter
1 foot/sec/sec = 0.3048 meter/sec/sec.

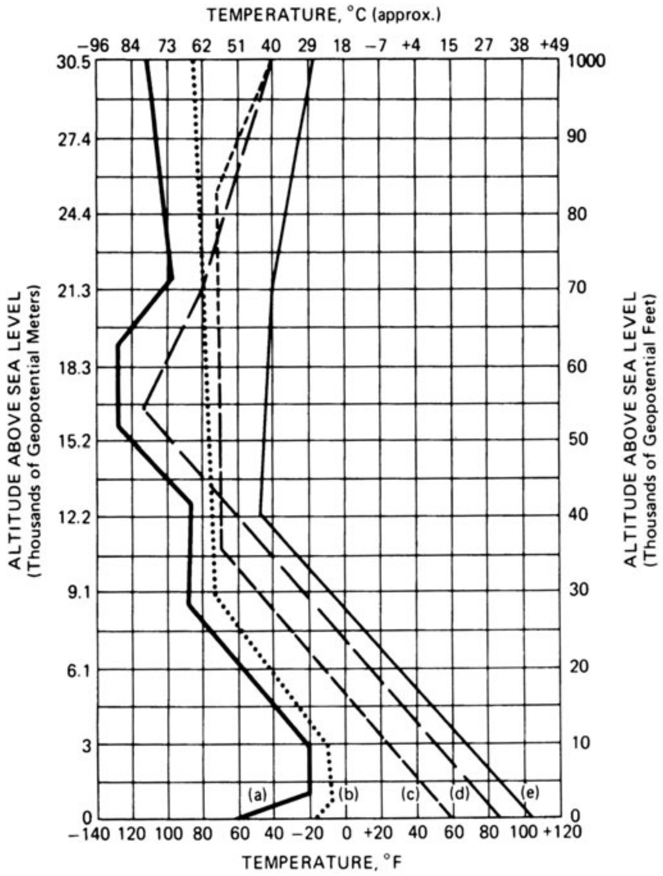


Fig. 1. Relationship between temperature and altitude: (a) cold and (e) hot are the composites of extremes of cold and hot atmospheres; (b) arctic and (d) tropical are the composites of the arctic and tropical regions: (e) is the standard atmosphere upon which altimetry is based

repeatedly absorbed and re-emitted, as the radiant energy progresses through the atmosphere. The downward flux, or counterradiation, is of basic importance in the so-called *greenhouse effect*; the upward flux is essential to the radiative balance of the planet.

Terrestrial radiation is defined as the total infrared radiation emitted from the earth's surface—to be carefully distinguished from atmospheric radiation, insolation, and effective terrestrial radiation, the latter being the difference between the outgoing infrared terrestrial radiation of the earth's surface and the downcoming infrared counterradiation from the atmosphere.

In meteorology, the cooling of the earth's surface and adjacent air, accomplished mainly at night, but whenever the earth's surface suffers a net loss of heat due to terrestrial radiation is known as *radiational cooling.* See also **Solar Energy.**

Thermodynamics of the Atmosphere

In meteorological calculations, the ideal gas law is a satisfactory approximation for the derivation of formulas for the mixture of gases that constitute the atmosphere. The derivation is:

$$PV = RT$$

where $V =$ volume
$P =$ pressure
$R =$ universal gas constant
$T =$ absolute temperature

For one gram, this becomes

$$PV = \frac{RT}{m}$$

where m is the molecular weight of the gas. For G grams, this becomes,

$$PV = \frac{GRT}{m}$$

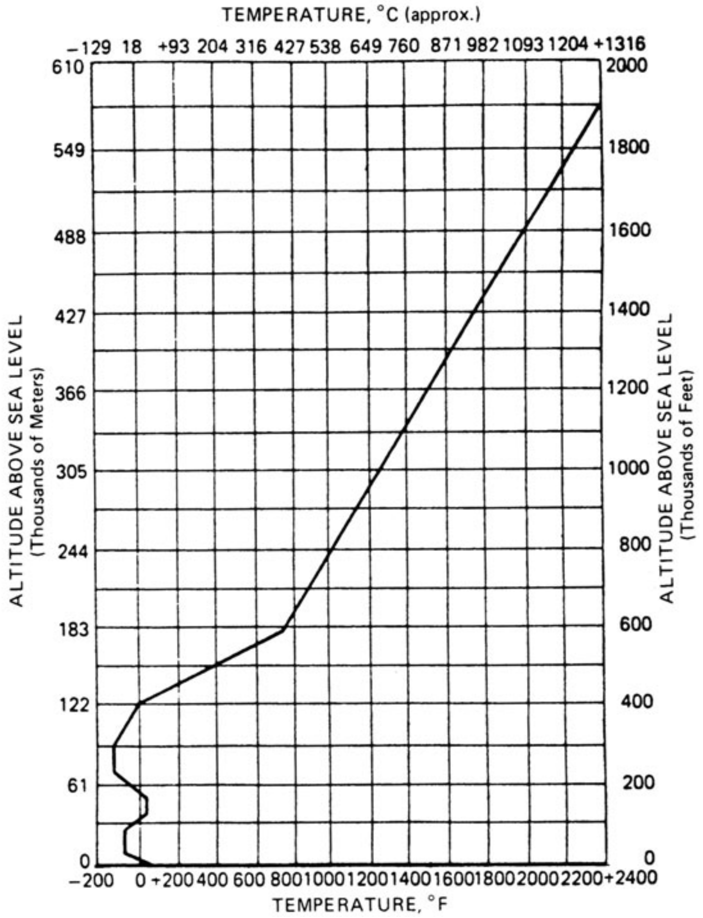


Fig. 2. Real kinetic temperature of the atmosphere, a measure of the kinetic energy of the molecules and atoms constituting the atmosphere is plotted against altitude above sea level here. Numerical values are determined by the assumed molecular weight of the air (see Table 4), as well as assumed values of the temperature lapse rate.

This equation is valid for each of the constituent gases of the atmosphere.

For nitrogen,

$$P_nV = \frac{G_nRT}{m_n}$$

For oxygen,

$$P_0V = \frac{G_0RT}{m_0}$$

For argon,

$$P_aV = \frac{G_aRT}{m_a}$$

For water vapor,

$$P_wV = \frac{G_wRT}{m_w}$$

When there is no water vapor present in the atmosphere, these equations can be combined as follows:

$$P_tV = (P_n + P_0 + P_a)V = RT\left[\frac{G_n}{m_n} + \frac{G_0}{m_0} + \frac{G_a}{m_a}\right]$$
$$= RT\left(\frac{G_t}{m_t}\right)$$

In these equations, P_t is the total pressure of the nitrogen, oxygen, and argon. Also, G_t is the total mass of the gases; and m_t is the molecular weight of the mixture, with a numerical value of 28.97.

Because water vapor is always present in varying quantities in the atmosphere, corrections in the equation of state must be made in accordance with the amount of water vapor present. Procedure is as follows:

$$PV = (P_t + P_w)V = RT\left[\frac{G_t}{m_t} + \frac{G_w}{m_w}\right]$$

$$= RT\left[\frac{G}{m_t} + \frac{G_w}{m_t}\ \frac{G_w}{m_w}\right]$$

In these equations, P is the total pressure of the air gases plus the water vapor, and G is the total mass of the air gases plus the water vapor.

This equation can be rearranged and simplified:

$$PV = RT\frac{G}{m_t}\left[1 - \frac{G_w}{G}\left(1 - \frac{m_t}{m_w}\right)\right]$$

where m_t has a value of 28.97 and m_w has a value of 18.00. This equation is easily reduced to

$$PV = RT\frac{G}{m_t}\left[1 + 0.6\frac{G_w}{G}\right]$$

Virtual temperature of the air is defined as

$$T' = T\left[1 + 0.6\frac{G_w}{G}\right]$$

Virtual temperature is, in effect, the temperature of a mass of dry air having the same density of another mass of air containing water vapor. Virtual temperature is always greater than real temperature, except when G_w is nil.

The equation of state for real air becomes

$$PV = \frac{RT'G}{m_t}$$

If R, the universal gas constant, is made into a specific gas constant for air by letting $R/m_t = R_a$, then for one gram of air, the equation of state becomes $PV = R_aT'$.

The hydrostatic state of equilibrium of the atmosphere varies with the type of atmosphere that is under consideration.

Standard atmosphere is a term used in the following references:

1. A hypothetical vertical distribution of atmosphere temperature, pressure, and density, which, by international agreement, is taken to be representative of atmosphere for purposes of pressure altimeter calibrations, aircraft performance calculations, aircraft and missile design, ballistic tables, etc. The air is assumed to obey the ideal gas law and the hydrostatic equation, which, taken together, relate temperature, pressure, and density variations in the vertical. It is further assumed that the air contains no water vapor, and that the acceleration of gravity does not change with height. The current standard atmosphere was adopted in 1952 by the International Civil Aeronautical Organization (ICAO) and supplants the U.S. Standard Atmosphere prepared in 1925. The parametric assumptions and physical constants used in preparing the ICAO Standard Atmosphere are as follows:

(a) Zero pressure altitude corresponds to that pressure which will support a column of mercury 760 mm high. This pressure is taken to be 1.013250×10^6 dynes/cm^2, or 1013.250 mb, or 101.325 kPa (and is known as one standard atmosphere or one atmosphere).
(b) The gas constant for dry air is 2.8704×10^6 erg/gm K.
(c) The ice point at one standard atmosphere pressure is 273.16 K.
(d) The acceleration of gravity is 980.665 cm/sec^2.
(e) The temperature at zero pressure altitude is 15°C or 288.16 K.
(f) The density at zero pressure altitude is 0.0012250 gm/cm^3.
(g) The lapse rate of temperature in the troposphere is 6.5°C/km.
(h) The pressure altitude of the tropopause is 11 km.
(i) The temperature at the tropopause is −56.5°C.

2. A standard unit of atmospheric pressure; the 45° atmosphere, defined as the pressure exerted by a 760 mm column of mercury at 45° latitude at sea level at temperature 0°C (acceleration of gravity = 980.616 cm/sec^2). One 45° atmosphere equals 760 mm Hg(45°); 29.9213 in. Hg(45°); 1013.200 mb; 101.325 kPa.

Ballistics standard artillery atmosphere is composed of a set of values describing atmospheric conditions on which ballistic computations are based, namely, zero wind, pressure of 1000 millibars at the ground, temperature of 15°C, relative humidity of 78%, and a lapse rate that yields a prescribed density-altitude relationship.

Adiabatic atmosphere is characterized by a dry-adiabatic lapse rate throughout its vertical extent.

Model atmosphere is a term used for any theoretical representation of the atmosphere, with particular reference to vertical temperature and pressure distribution.

Isothermal atmosphere (or exponential atmosphere) is an atmosphere in hydrostatic equilibrium, in which the temperature is constant with height, and in which, therefore, the pressure decreases exponentially upward. In such an atmosphere, the thickness between any two levels is given by

$$Z_B - Z_A = \frac{R_dT_v}{g}\ln\frac{P_A}{P_B}$$

where R_d is the gas constant for dry air, T_v the virtual temperature (°K), g the acceleration of gravity, and P_A and P_B the pressures at the heights Z_A and Z_B, respectively. In the isothermal atmosphere, there is no finite level at which the pressure vanishes.

Polytropic atmosphere is characterized by hydrostatic equilibrium with a constant nonzero lapse rate. The vertical distribution of pressure and temperature is given by

$$\frac{p}{p_0} = \left(\frac{T}{T_0}\right)^{g/Ry}$$

where p is the pressure, T the Kelvin temperature, g the acceleration of gravity, R the gas constant for air, and y the environmental lapse rate, the subscript zeros denoting values at the earth's surface.

Homogeneous atmosphere is a hypothetical atmosphere in which the density is constant with height. The lapse rate of temperature in such an atmosphere is known as the autoconvective lapse rate and is equal to g/R (or approximately 3.4°C/100 meters), where g is the acceleration of gravity and R is the gas constant for air. A homogeneous atmosphere has a finite total thickness given by R_dT_v/g, where R_d is the gas constant for dry air and T_v is the virtual temperature (°K) at the surface. For a surface temperature of 273°K, the vertical extent of the homogeneous atmosphere is approximately 8000 meters. At the top of such an atmosphere, both the pressure and absolute temperature vanish.

With respect to radio propagation, a homogeneous atmosphere is one that has a constant index of refraction, or one in which radio waves travel in straight lines at constant speed. The ideal "homogeneous atmosphere" in this sense is *free space*, which is a perfectly homogeneous medium possessing a dielectric constant of unity, and in which, as in a perfect vacuum, there is nothing to reflect, refract, or absorb energy.

Thermotropic atmosphere, a term used in numerical weather forecasting, is an atmosphere in which the parameters to be forecast are the height of one constant-pressure surface (usually 500 millibars) and one temperature (usually the mean temperature between 1000 and 500 millibars) whereby a surface prognostic chart can also be constructed.

Equivalent barotropic atmosphere is one in which the wind does not change direction with altitude and, therefore, the isotherms and isobars are everywhere parallel.

Barotropic atmosphere is one of a number in which some of the following conditions exist: (1) pressure and temperature surfaces coincide; (2) zero vertical wind shear; (3) zero vertical motion; and (4) zero horizontal velocity divergence.

Baroclinic atmosphere is one in which constant-pressure surfaces intersect constant-density surfaces, thereby creating solenoids which can cause acceleration.

Adiabatic Processes in the Atmosphere

An adiabatic process is a thermodynamic change of state of a system, in which there is no transfer of heat or mass across the boundaries of the system; where compression always results in warming, and expansion in cooling. When a parcel of air is moved from one position to another in such a manner that energy does not flow across the boundaries of the parcel, the thermal changes taking place within the parcel are said to be adiabatic changes.

Dry-adiabatic processes, during which the air involved remains unsaturated, are relatively simple. The first law of thermodynamics applied to a parcel of unsaturated air of unit mass stipulates:

$$dq \times c_v\, dT + Ap\, dv$$

which, when combined with the gas equation becomes

$$dq = (c_v + AR)\, dT - \frac{ART}{p}\, dp$$

For the adiabatic process, this becomes

$$\frac{dT}{T} = \frac{AR}{c_p}\frac{dp}{p}$$

which, upon integration, becomes

$$\frac{T}{T_0} = \left(\frac{p}{p_0}\right)^{AR/c_p}$$

Dry-adiabatic horizontal transfer of a parcel from higher to lower or lower to higher pressure is of only minor consequence because of the comparatively small magnitude of pressure change. Dry-adiabatic vertical transfer of a parcel, however, is one of the important meteorological processes. Temperature decrease in a rising, and increase in a sinking, parcel amounts to very nearly 9.8°C per km, or 5.4°F per 1000 feet (304.8 meters). Dew-point changes in a vertically moving unsaturated parcel are considerably less. The dew-point decreases in rising air, and increases in sinking air, at a rate of between 1.3 and 1.8°C per kilometer (0.7 and 1.0°F per 1000 feet), depending upon air temperature.

Pseudo-adiabatic (or *saturation-* or *moist-adiabatic*) **processes** involve condensation or evaporation, and are by no means constant or simple. In a parcel rising pseudo-adiabatically, the temperature decrease is always less than the dry-adiabatic temperature change by an amount depending upon the weight of the water being condensed and the temperature at which condensation occurs. Condensation releases the latent heat of varporization within the parcel, which partially counteracts dry-adiabatic cooling. The rate of cooling in rising saturated air varies from about 2.7°C per kilometer (1.5°F per 1000 feet) in warm air at sea level to 9.7°C per kilometer (5.3°F per 1000 feet) at high altitudes in cold air, a range that is the direct result of the variance in the amount of water resident in a given mass of air at full saturation. Very cold air can retain only a slight amount of water, whereas very warm air can hold relatively large quantities. Values of resident water vapor at saturation range from 0.01% by weight in arctic air to 3% by weight in tropical air.

Sinking saturated air remains saturated only for a comparatively short distance, during which it is heated pseudo-adiabatically at a rate determined by the amount of evaporation occurring within the parcel. As soon as it becomes unsaturated, the sinking parcel descends dry-adiabatically. Foehn winds are examples of both pseudo-adiabatic and dry-adiabatic changes. Air flowing uphill is cooled pseudo-adiabatically until it reaches the hilltop. On the lee side, the air descends dry-adiabatically. Observable results of the true foehn wind are abundant clouds and rain or snow on the windward side of a mountain range, and clear, warm air on the lee side.

Dew-points in saturated air rising pseudo-adiabatically decrease at the same rate as the temperature. Dew-points in saturated sinking air increase at the same rate as the temperature until the air parcel is no longer saturated; then they rise slowly, as previously described in connection with sinking saturated air.

A large percentage of all clouds and nearly all precipitation result from adiabatic ascent of air.

Assuming increasing positive values with altitude, the following relations hold:

1. Dry-adiabatic temperature change with altitude:

$$\frac{\partial t}{\partial h} = \frac{gK}{R} = -9.8°\text{C/km}$$

where t = temperature of parcel
g = gravitational constant
$K = \frac{c_p - c_v}{c_p} = .288$
R = gas constant

2. Dry-adiabatic dew-point change with altitude:

$$\frac{\partial t_d}{\partial h} = -1.71\left[1 + \frac{2t_d}{237.3} - \frac{t}{273}\right]°\text{C/km}$$

where t_d = dew-point temperature in °C
t = air temperature in °C

3. Pseudo-adiabatic temperature change with altitude:

$$\frac{\partial t}{\partial h} = -g\left(\frac{A + .621\frac{e}{p}\frac{L}{Rt}}{c_p + .621\frac{L}{P}\frac{de}{dt}}\right)\text{C/km}$$

where t = temperature of the parcel in °C
g = gravitational constant
A = heat equivalent of work
e = water vapor pressure
p = air pressure
L = heat of condensation
c_p = specific heat at constant pressure for air

Virtual Temperature of Air. In meteorological calculations, it is often convenient to use, instead of the actual air temperature, the temperature which a parcel of air would have if it had the same density and pressure as the sample in question, but was entirely free from water vapor. Since dry air is denser than water vapor under the same conditions, removal of water vapor from moist air will increase its density, so that the temperature will need to be raised to obtain an equivalent density. Therefore, the virtual temperature is higher than the actual, and is given by:

$$T_v = (1 + 0.61q)T$$

where T is the absolute temperature and q the specific humidity.

Atmospheric Stability, Instability, and Equilibrium

Everywhere that air is in motion, some vertical perturbations are present. Isolated parcels and currents of air are started upwards or downwards into new environments. If the density of the environment is different from the density of the parcel after any modification caused by the change of pressure, the parcel experiences a force of buoyancy which may accelerate or retard the initial displacement. The criterion for static stability of a horizontally stratified compressible fluid is that the gradient of potential density should be negative upwards.

Stability. In meteorology, *static stability* (also called *hydrostatic stability, vertical stability*, or *convectional stability*) is the stability of an atmosphere in hydrostatic equilibrium with respect to vertical displacements, usually considered by the parcel method. The criterion for stability is that the displaced parcel be subjected to a buoyant force opposite to its displacement, e.g., that a parcel displaced upward be colder than its new environment. This is the case if $\gamma < \Gamma$, where γ is the environmental lapse rate and Γ is the process lapse rate, dry-adiabatic for unsaturated air and saturation-adiabatic for saturated air.

Neutral stability (also called *indifferent stability* or *indifferent equilibrium*) is the state of an unsaturated or saturated column of air in the atmosphere when its environmental lapse rate is equal to the dry-adiabatic lapse rate or the saturation-adiabatic lapse rate, respectively. Under such conditions, a parcel of air displaced vertically will experience no buoyant acceleration.

Hydrostatic Equilibrium. The state of a fluid whose surfaces of constant pressure and constant mass (or density) coincide and are horizontal throughout. Complete balance exists between the force of gravity and the pressure force. The relation between the pressure and the geometric height is given by the hydrostatic equation. The analysis of atmospheric stability has been developed most completely for an atmosphere in hydrostatic equilibrium. The hydrostatic equation is the form assumed by the vertical component of the vector equation of fluid motion when Coriolis, earth curvature, frictional, and vertical acceleration terms are considered negligible compared with those involving the vertical pressure force and the force of gravity. Thus,

$$\frac{\partial p}{\partial z} = -\rho g$$

where p is the pressure, ρ the density, g the acceleration of gravity, and z the geometric height.

Instability. The concept of instability is employed in many sciences. It is, in general, a property of the steady state of a system such that certain disturbances or perturbations introduced into the steady state will increase in magnitude, the maximum perturbation amplitude always remaining larger than the initial amplitude. The method of small perturbations, assuming permanent waves, is the usual method of testing for instability; unstable perturbations then usually increase exponentially with time. In meteorology, the small perturbations, may be a wave or a parcel displacement. The parcel method assumes that the environment is unaffected by the displacement of the parcel. The slice method has occasionally been used as a modification of the parcel method to gain a little information about the interaction of parcel and environment.

Absolute instability is the state of a column of air in the atmosphere when it has a superadiabatic lapse rate, i.e., greater than the dry-adiabatic lapse rate. An air parcel displaced vertically would be accelerated in the direction of the displacement. The kinetic energy of the parcel would consequently increase with the increasing distance from its level of origin.

Baroclinic instability arises from the existence of a meridional temperature gradient (and hence of a thermal wind) in an atmosphere in quasigeostrophic equilibrium and possessing static stability.

Barotropic instability arises from certain distributions of vorticity in a two-dimensional nondivergent flow. This is an *inertial instability* in that kinetic energy is the only form of energy transferred between current and perturbation. The variation of vorticity, i.e., shear, in the basic current may be concentrated in discontinuities of the horizontal wind shear (to be distinguished from *Helmholtz instability*, where the velocity itself is discontinuous) or may be continuously distributed in a curved velocity profile. A well-known necessary condition for barotropic instability is that the vorticity must change sign, i.e., vanish, at a point of maximum shear.

Colloidal instability is a property attributed to clouds (regarded in analogy to colloidal systems or aerosols) by virtue of which the particles of the cloud tend to aggregate into masses large enough to precipitate.

Conditional instability is the state of a column of air in the atmosphere when its lapse rate is less than the dry-adiabatic lapse rate but greater than the saturation-adiabatic lapse rate. With reference to the vertical displacement of an air parcel, the air will be unstable if saturated and stable if unsaturated.

Convective instability (or *potential instability*) is the state of an unsaturated layer or column of air in the atmosphere whose wet-bulb potential temperature, or equivalent potential temperature decreases with elevation.

Gravitational instability occurs in a system in which buoyancy or reduced gravity is the only restoring force on displacements.

Helmholtz instability (also called *shearing instability*) arises from a shear, or discontinuity, in current speed at the interface between two fluids in two-dimensional motion. The perturbation gains kinetic energy at the expense of that of the basic currents.

Hydrodynamic instability (or *dynamic instability*) refers to instability of parcel displacements or, more usually, of waves in a moving fluid system governed by the fundamental equations of hydrodynamics. The space scale of unstable waves is important in meteorology; thus Helmholtz, baroclinic, and barotropic instability give, in general, unstable waves of increasing length. The time scale is also important; a perturbation that grows for two days before dying out is effectively unstable for many meteorological purposes, but this is an initial value problem, and one cannot assume the existence of permanent waves. These meteorological types of hydrodynamic instability must not be confused with the phenomenon often referred to by mathematicians and physicists by the same term. A great deal of study has been devoted to the problem of the onset of turbulence in simple flows under laboratory conditions, and here viscosity is a source of instability. This is not the case in any meteorological motion yet investigated.

Inertial instability (or *dynamic instability*) is, generally, instability in which the only form of energy transferred between the steady state and the disturbance is kinetic energy. More specifically, it is the instability arising in a rotating fluid mass when the velocity distribution is such that the kinetic energy of a disturbance grows at the expense of kinetic energy of the rotation.

Rotational instability is usually synonymous with inertial instability, being, in general, any instability of a rotating fluid system.

Static instability (or *hydrostatic instability*) refers to instability of vertical displacements of a parcel in a fluid in hydrostatic equilibrium.

Thermal instability results in free convection in a fluid heated at a boundary. For the case of heating from below, the onset of convection (as opposed to conduction) is determined by a critical value of the Rayleigh number, and the linear theory admits of various convection-cell forms, including the hexagonal Benard cell. The theories of thermal instability, which are represented by sixth-order convection equations and take into account both viscosity and conductivity, must be distinguished from the theory of static instability, based on the oscillations of a parcel in an atmosphere in hydrostatic equilibrium.

Lapse Rate. This is the rate at which temperature decreases or lapses with altitude; the vertical temperature gradient. Since temperature normally decreases with altitude in the troposphere, it is convenient to assign positive values to the rate of temperature change with altitude: Lapse rate, therefore, is defined as the rate of change of temperature with altitude, and is positive when the temperature decreases. The term applies ambiguously to the environmental lapse rate and the process lapse rate (defined below), and the meaning must often be ascertained from the text.

Autoconvective lapse rate (or *autoconvection gradient*). The environmental lapse rate of temperature in an atmosphere in which the density is constant with height (homogeneous atmosphere), equal to g/R, where g is the acceleration of gravity and R the gas constant. For dry air, the autoconvective lapse rate is approximately $+3.4 \times 10^{-4}$ °C per centimeter.

Undisturbed air will remain stratified even though the lapse rate exceeds the adiabatic rate of 9.8°C per kilometer (5.5°F per 1000 feet). If, however, the lapse rate becomes sufficiently large, density of the air will increase with altitude and will overturn. This critical lapse rate is 34.17°C per kilometer (or nearly 19°F per 1000 feet). Dust devils and whirlwinds result from this steep lapse rate, which occurs at ground levels, particularly over concrete roads and the sand and rock of deserts during the heat of day.

Dry-adiabatic lapse rate. A special process lapse rate (defined below) of temperature, being the rate of decrease of temperature with height of a parcel of dry air lifted adiabatically through an atmosphere in hydrostatic equilibrium. This lapse rate is g/C_{pd}, where g is the acceleration of gravity and C_{pd} is the specific heat of dry air at constant pressure; numerically equal to 9.767°C per kilometer (about 5.4°F per 1000 feet). Potential temperature is constant with height in an atmosphere with this lapse rate.

Environmental lapse rate. The rate of decrease of temperature with elevation, $-\partial T/\partial z$, or occasionally, $\partial T/\partial p$, where p is pressure. The concept may be applied to other atmospheric variables (e.g., lapse rate of density) if these are specified. The environmental lapse rate is determined by the distribution of temperature in the vertical at a given time and place, and should be carefully distinguished from the process lapse rate (defined below), which applies to an individual air parcel.

Process lapse rate. The rate of decrease of the temperature of an air parcel as it is lifted, $-dT/dz$, or occasionally dT/dp, where p is pressure. The concept may be applied to other atmospheric variables, e.g., the process lapse rate of density. The process lapse rate is determined by the character of the fluid processes and should be carefully distinguished from the environmental lapse rate, which is determined by the distribution of temperature in space. In the atmosphere, the process lapse rate is usually assumed to be either the dry-adiabatic lapse rate or the saturation-adiabatic lapse rate (defined below).

Saturation-adiabatic lapse rate (or *moist-adiabatic lapse rate*). A special case of process lapse rate, defined as the rate of decrease of temperature with height of an air parcel lifted in a saturation-adiabatic process through an atmosphere in hydrostatic equilibrium. Owing to the release of latent heat, this lapse rate is less than the dry-adiabatic lapse rate, and the differential equation representing the process must be integrated numerically. Wet-bulb potential temperature is constant with height in an atmosphere with this lapse rate.

Superadiabatic lapse rate. An environmental lapse rate greater than the dry-adiabatic lapse rate, such that potential temperature decreases with height.

Atmospheric Inversion. This is the abnormal condition in which the temperature of the atmosphere increases with height. The term is also used for a level at which the vertical gradient of temperature changes sign. Dynamically, more importance attaches to the vertical distribution of potential temperature or potential density, and it is common to refer to the boundary between a lower region of negative gradient of potential density and an upper one of positive gradient as an inversion, whether or not the temperature gradient changes sign. An example is the tropopause.

In meteorology, *temperature inversion* refers to an atmospheric condition in which temperature increases with altitude. The temperature below the stratosphere normally decreases with altitude; thus, when it increases, normal conditions are inverted, and an inversion has occurred. Inversions in the troposphere are usually restricted to shallow layers of air, which most frequently occur in the lower 5000 feet (1524 meters) above the surface. In low latitudes, the stratosphere has a slight inversion more or less permanently.

The principal characteristics of an inversion layer is its marked static stability, so that very little turbulent exchange can occur within it. Strong wind shears often occur across inversion layers, and abrupt changes in concentrations of atmospheric particulates and atmospheric water vapor may be encountered on ascending through the inversion. When, in meteorological literature and discussion, an "inversion" is mentioned, a temperature inversion is usually meant. The following are particular types of temperature inversion:

Frontal inversion is encountered upon vertical ascent through a sloping front (or frontal zone).

Subsidence inversion is produced by the adiabatic warming of a layer of subsiding air and is enhanced by vertical mixing in the air layer below the inversion.

Surface inversion (or *ground inversion*) is a temperature inversion based at the earth's surface; that is, an increase of temperature with height beginning at the ground level. This condition is due primarily to greater radiative loss of heat at and near the surface than at levels above. Thus, surface inversions are common over land prior to sunrise and, in winter, over high-latitude continental interiors.

Trade-wind inversion (or *trade inversion*) is usually present in the trade-wind streams over the eastern portions of the tropical oceans. It is formed by broad-scale subsidence of air from high altitudes in the eastern extremities of the subtropical highs. While descending, the current meets the opposition of the low-level maritime air flowing equatorward. The inversion forms at the meeting point of these two strata, which flow horizontally in the same direction.

Circulation in the Atmosphere

The worldwide pattern of air movement is related to the global pressure belts, the rotation of the earth, the distribution of temperature over the earth, friction between the earth and the atmosphere, and the location of mountains and oceans. The major large-scale patterns of movement of the atmosphere include:

1. *General circulation*, which in the broadest sense is a complete description of atmospheric motions over the earth. These data are generated from the day-to-day patterns of flow and describe temporal as well as the mean spatial conditions, plus their variability in time and space.

2. *Planetary Circulation.* Refers specifically to (a) the system of large-scale disturbances in the troposphere when viewed on a hemispheric or world-wide scale, and (b) to the mean or time-averaged hemispheric circulation of the atmosphere; in this sense, almost synonymous with general circulation.

3. *Primary Circulation.* The prevailing fundamental atmospheric circulation on a planetary scale that must exist in response to (a) radiation differences with latitude, (b) the rotation of the earth, (c) the particular distribution of land and ocean; and which is required from the viewpoint of conservation of energy. Primary circulation and general circulation are sometimes taken synonymously. They may be distinguished, however, on the basis of approach. That is, primary circulation is the basic system of winds, of which the secondary and tertiary circulations are perturbations; while general circulation encompasses at least the secondary circulations. The latter dimension of circulation has features of cyclonic scale, while tertiary circulation is represented by such phenomena as local winds, thunderstorms, and tornadoes.

4. *Meridional Cell.* A very large-scale convection circulation in the atmosphere or ocean that takes place in a meridional plane, with northward and southward currents in opposite branches of the cell, and upward and downward motion in the equatorward and poleward ends of the cell.

5. *Zonal Flow* (or *Zonal Circulation*). The flow of air along a latitude circle; more specifically, the latitudinal (east or west) component of existing flow.

6. *Polar Vortex.* A large-scale cyclonic circulation, centered generally in the polar regions. Specifically, the vortex has two centers in the mean, one near Baffin Island and another over northwest Siberia. In the southern hemisphere, there is one center near the South Pole.

7. *Westerlies.* Specifically, the dominant west-to-east motion of the atmosphere, centered over the middle latitudes of both hemispheres. Generally, any winds with components from the west.

8. *Easterlies.* Any winds with components from the east, usually applied to broad currents or patterns of persistent easterly winds.

Forces in the Atmosphere

There are two forces that account mainly for driving the horizontal flow of air on a global scale:

1. *Coriolis Force.* This is related to the rotation of the earth and is expressed vectorially by

$$C = -2\Omega \times \mathrm{v}$$

and quantitatively by

$$C = -2\Omega v \sin \phi$$

where C = Coriolis force
Ω = angular velocity of the earth
v = velocity vector
v = speed of wind
ϕ = the latitude

Coriolis force acts at right angles to the direction of the wind. In the northern hemisphere, it acts toward the right; in the southern hemisphere, toward the left.

2. *Pressure Gradient Force.* This is related to the field of atmospheric pressure at a specified level at or above sea level. It is expressed vectorially by

$$P_f = -\alpha \nabla_p$$

and quantitatively by

$$P_f = -\alpha \frac{\partial p}{\partial l}$$

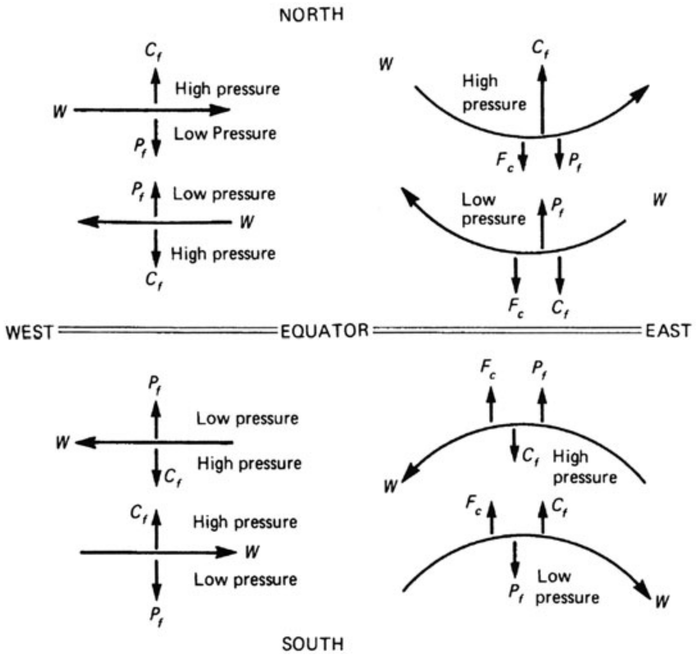
where P_f = force resulting from the pressure gradient
α = specific volume
p = pressure gradient
l = rate of pressure change along a direction.

In meteorology, the pressure gradient is regarded as acting from high to low pressure. It is considered to be the component of the force acting at right angles to the isobars. The component parallel to isobars is zero.

Basic flow patterns in the atmosphere on the horizontal are indicated in Fig. 3. These patterns are in a steady-state (no acceleration).

3. *Other Forces.* Additional forces acting on the atmosphere include:

(a) Centripetal acceleration related forces associated with curved flow.
(b) Frictional drag related forces acting near and at the surface of the earth.



LEGEND:
W = wind direction (straight line and curved)
C_f = direction of Coriolis force
P_f = direction of pressure gradient force
F_c = direction of force associated with centripetal acceleration in the case of curved flow

Fig. 3. Basic flow patterns in the atmosphere on the horizontal. These patterns are in a steady state (no acceleration).

(c) Isoallobaric forces related to the time rate of change of pressure field.
(d) Divergence and convergence related to vertical flow.

Pressure-Gradient Forces

Pressure Belts. Surrounding the earth at its surface and directly related to the generalized pattern of winds in the atmosphere are four alternating belts of high and low pressure. These belts are formed correspondingly in both the northern and southern hemispheres at roughly 30° intervals from equator to poles. They shift with season, on an average of 5° latitude, reaching the most northerly position in late summer; the most southerly in late winter. See Fig. 4.

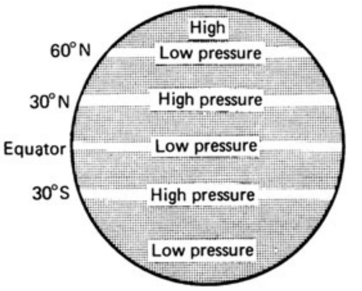


Fig. 4. Pressure belts and general circulation pattern of the air.

1. The *equatorial trough* is a quasicontinuous belt of low pressure extending north and south from the equator; it is commonly called the *doldrums*, or the *equatorial calms*, especially with reference to its light and variable winds. This entire region is one of very homogeneous air, probably the most ideally barotropic region of the atmosphere. Yet, humidity is so high that slight variations in stability cause major variations in weather. The position of the equatorial trough is fairly constant in the eastern portions of the Atlantic and Pacific; but it varies greatly with season in the western portions of those oceans and in southern Asia and the Indian Ocean. It moves into or toward the southern hemisphere.

2. The *horse latitudes* are narrow high-pressure belts over the oceans at approximately 30°–35°N and S, where winds are predominantly calm or very light, and weather is hot and dry. They are known in the northern and southern hemispheres, respectively, as the *calms of Cancer* and the *calms of Capricorn*. These latitudes mark the normal axis of the subtropical highs, and move north and south by about 5°, following the sun. In the North Atlantic Ocean, these are the latitudes of the Sargasso Sea, where surface waters converge, and which is characterized by clear, warm water, a deep blue color, and large quantities of floating Sargassum or "gulf weed." The name of the horse latitudes is believed to have originated in the days of sailing ships, when the voyage across the Atlantic in those latitudes was often prolonged by calms or baffling winds, so that water ran short, and ships carrying horses to the West Indies found it necessary to throw the horses overboard.

3. The *subpolar low-pressure belt* is located, in the mean, between 50° and 70° latitude. In the northern hemisphere, this "belt" consists of the Aleutian low and the Icelandic low. In the southern hemisphere, it is supposed to exist around the periphery of the Antarctic continent.

4. Areas of high pressure, the *polar highs*, form at the 90° poles, where the weather is violent and stormy.

Pressure Areas. These are areas within which the atmospheric pressure is either greater or smaller than other environing regions at the same altitude above sea level. In the case where the pressure is greater than other environing regions, the area of higher pressure is called a *high*. Highs are associated with anticyclonic circulation, clockwise in the northern hemisphere and counterclockwise in the southern hemisphere. In the case where the pressure is lower than other environing regions, the area of low pressure is called a *low*. Lows are associated with cyclonic circulation, counterclockwise in the northern hemisphere and clockwise in the southern hemisphere.

Primary and Secondary Highs and Lows. A further classification of pressure areas distinguishes between primary and secondary highs and lows.

Primary (or *semipermanent*) *highs and lows* cover large areas of the earth's surface for long periods of time (sometimes for the entire year). They are the result of unequal heating of the earth's surface and the consequent movements of air. Where air rises over warmer regions, lows are likely to form; where air sinks over cooler regions, highs are likely to form. The term *center of action* refers to any one of the primary highs or lows. Fluctuations in the nature of these centers are intimately associated with relatively widespread and long-term weather changes.

Secondary highs and lows are, respectively, anticyclonic and cyclonic movements that form within the primary highs and lows. The secondary cyclonic lows are represented, generally, by the inclement weather and more-or-less violent phenomena that accompany storms. The secondary highs, unlike lows, represent a single anticyclonic air mass.

The principal semipermanent pressure areas of the northern hemisphere are:

Bermuda high: Located over the North Atlantic Ocean, and so named especially when it is located in the western part of the ocean. This same subtropical high, when displaced toward the eastern part of the Atlantic, is known as the *Azores high*. On mean charts of sea level, it is a principal center of action. When it is well-developed and extends westward, warm and humid conditions prevail over the eastern United States, particularly in summer.

North American high: Covers most of North America during winter. This relatively weak high-pressure system is not nearly so well-defined as the analogous Siberian high.

Pacific high: Located over the North Pacific Ocean and centered, in the mean, at 30–40°N and 140–150°W. On mean charts of sea-level pressure, this subtropical high is a principal center of action.

Siberian high: Forms over Siberia in winter, and is particularly apparent on mean charts of sea-level pressure. It is enhanced by surrounding mountains, which prevent the cold air from flowing away readily. In summer, the Siberian high is replaced by a low-pressure area.

Subtropical highs: Form the subtropical high-pressure belt, and include the Bermuda (and Azores) and Pacific highs.

Subpolar highs: Form over the cold continental surfaces of subpolar latitudes, principally in northern hemisphere winter. These highs typically migrate eastward and southward.

Aleutian low: Located near the Aleutian Islands on mean charts of sea-level pressure, and represents one of the main centers of action in the atmospheric circulation of the northern hemisphere. It is most intense in the winter months; in summer, it is displaced toward the North Pole and is almost nonexistent. The traveling cyclones of subpolar latitudes usually reach maximum intensity in the area of the Aleutian low.

Icelandic low: Located near Iceland, mainly between Iceland and southern Greenland, on mean charts of sea-level pressure, and is a principal center of action in the atmospheric circulation of the northern hemisphere. It is most intense during winter; in summer, it not only weakens, but also tends to split into two centers, one near Davis Strait and the other west of Iceland.

Subpolar low-pressure belt: Located, in the mean, between 50° and 70° latitude. In the northern hemisphere, this "belt" consists of the Aleutian low and the Icelandic low. In the southern hemisphere, it is supposed to exist around the periphery of the Antarctic continent.

In the southern hemisphere, there are three semipermanent high pressure centers, one each in the three oceans, the Pacific, Atlantic, and Indian Ocean. These centers are near 30°S in all cases and they do not migrate much between winter and summer. There is one semipermanent low pressure center in the southern hemisphere located over the Antarctic region. The main feature in the southern hemisphere is a zone of relatively strong westerly winds between the semipermanent high cells and the semipermanent low cells. Except for the continent of Australia, these prevailing Westerlies blow uninterrupted over ocean waters.

Types of Pressure Areas. Commonly used general terms for designating various types of pressure areas include:

Center of action. Any one of several large areas of high and low barometric pressure changing little in location, and persisting through a season or through the whole year. Changes in the intensity and positions of these pressure systems are associated with widespread weather changes. The term is also used to describe any region in which the variation of any meteorological element is related to weather of the following season in other regions.

Col. A relatively small area about midway between two cyclones and two anticyclones where the pressure gradient is very weak and winds are usually light and variable. It is the point of intersection between a trough and a ridge in the pressure pattern of a weather map; the point of relatively lowest pressure between the two highs and the point of relatively highest pressure between two lows.

Depression. An area of low pressure; a low or a trough. This is usually applied to a certain stage in the development of a tropical cyclone, to migratory lows and troughs, and to upper-level lows and troughs that are only weakly developed.

High. An area of high pressure, referring to a maximum of atmospheric pressure in two dimensions (closed isobars) in the synoptic surface chart, or a maximum of height (closed contours) in the constant-pressure chart. Since a high, on the synoptic chart, is always associated with anticyclonic circulation, the term can be used interchangeably with anticyclone.

Low. Also sometimes called depression. A low is an area of low pressure, referring to a minimum of atmospheric pressure in two dimensions (closed isobars) on a constant-height chart, or a minimum of height (closed contours) on a constant-pressure chart. Since a low, on a synoptic chart, is always associated with cyclonic circulation, the term can be used interchangeably with cyclone.

Ridge. Also sometimes called *wedge*, an elongated area of relatively high atmospheric pressure, almost always associated with and most clearly identified as an area of maximum anticyclonic curvature of wind flow. The locus of this maximum curvature is called the *ridge line*.

The most common use of this term is to distinguish it from the closed circulation of a high (or anticyclone); but a ridge may include a high (and an upper-air ridge may be associated with a surface high), and a high may have one or more distinct ridges radiating from its center. The opposite of a ridge is a trough.

Trough. An elongated area of relatively low atmospheric pressure. The axis of a trough is the *trough line*, along which the isobars are symmetrical and curved cyclonically. A V-shaped trough normally contains a front; a U-shaped trough generally contains no front or a very weak one. Usually there is considerable weather associated with a trough line of the V variety. A large-scale trough may include one or more lows; an upper-air trough may be associated with a lower-level low; and a low may have one or more distinct troughs radiating from it. Trough-line movements can be computed and a forecast made of future positions.

Isallobars and Isallobaric Fields. Atmospheric pressure changes at every point in the atmosphere from time to time. It is possible to measure such changes with considerable accuracy. The unit to express pressure change (or pressure tendency) is conventionally the total net change occurring in a 3-hour interval. It is customary to indicate the nature of the change, because the pressure change character may have varied during the selected time interval.

Three-hourly pressure changes are plotted on a synoptic chart (weather map) and lines drawn to join points of equal pressure change. Care is exercised, however, in noting the character of the change in judging the real value of Δp. Lines joining points of equal pressure change are isallobars. See Fig. 5. Isallobars, taken together, constitute an isallobaric field. Where there is present an isallobaric field superimposed on a pressure field, a component of the actual wind blows along the isallobaric gradient, which is directed perpendicular to the isallobars toward regions of greatest pressure fall or least pressure rise, as the case may be. Normally, the isallobaric wind component is small unless the isallobaric field is pronounced. When there is no isallobaric field, the wind is defined approximately by the orientation and spacing of isobars themselves. Isallobaric fields also are used to compute the movement of pressure areas, ridges, troughs, cols, fronts, and isobars. Centers of low pressure will tend to move toward the region of greatest pressure fall, whereas centers of high pressure tend to move toward regions of maximum pressure rise. Both cases have modifying factors, and the direction is not always exactly as would be expected from a casual glance. In all cases of computation of movement, it is necessary to know the isallobaric field, the pressure field, and the orientation of the system whose movement is to be computed.

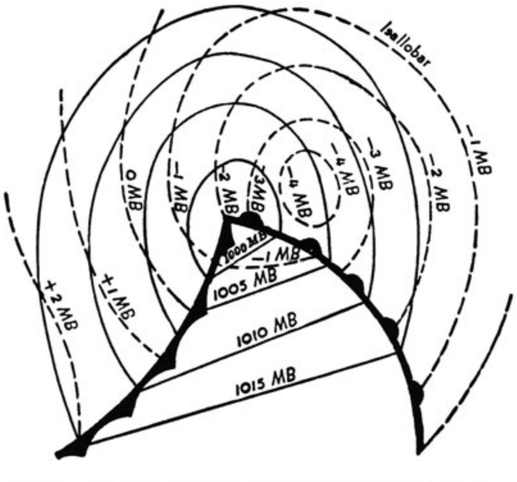


Fig. 5. Isallobars in isallobaric field of wave cyclone.

In establishing an isallobaric field, it is imperative that the diurnal pressure change be accounted for and a correction applied. Diurnal pressure changes do affect movement of pressure systems, but the effect is transient, and, therefore, must be neglected in any study extending more than 6 hours into the future.

In general, isallobars and isallobaric fields are as useful in weather forecasting as the pressure field itself.

Tendency is defined as the local rate of change of a vector or scalar quantity with time at a given point in space. Thus, in symbols, $\partial p/\partial t$ is the pressure tendency, $\partial \xi/\partial t$ the vorticity tendency, etc. Because of the difficulty of measuring instantaneous variations in the atmosphere, variations are usually obtained from the difference in magnitudes over a finite period of time; and the definition of tendency is frequently broadened to include the local time variations so obtained. An example is the familiar three hourly pressure tendency given in surface weather observations.

Terms Associated with Circulation of the Atmosphere

Advection. The process of transport of an atmospheric property solely by the mass motion (velocity field) of the atmosphere; also, the rate of change of the value of the advected property at a given point. Fog drifts from one place to another and cold air moves from polar regions southward by advection. In synoptic meteorology, the term refers only to the horizontal or isobaric components of motion, that is, the wind field as shown on a synoptic chart. The distinction is made between advection and convection, the former describing the predominantly horizontal, large-scale motions of the atmosphere, while the latter describes the predominantly vertical, locally induced motions. Large-scale north-south advection is more prominent in the northern than in the southern hemisphere, but west-to-east advection is prominent on both sides of the equator.

Air Parcel. An imaginary body of air to which may be assigned any or all of the basic dynamic and thermodynamic properties of atmospheric air. A parcel is large enough to contain a very great number of molecules, but small enough so that the properties assigned to it are approximately uniform within it and so that its motions with respect to the surrounding atmosphere do not induce marked compensatory movements. It cannot be given precise numerical definition, but a cubic foot of air might fit well into most contexts where air parcels are discussed, particularly those related to static stability.

Air-Parcel Trajectory. A parcel of air located in a given pressure field will move with the gradient wind of the field (assuming steady flow). At the end of a few hours, the parcel will locate in some new region where it has been carried by the wind. If, however, the pressure field, and therefore the wind, is changing, the parcel will not move into a position indicated by the existing gradient flow. It will follow a trajectory or path dictated by successive gradient directions and velocities as indicated by synoptic charts. An approximation to its trajectory can be had by extrapolating the parcel's indicated movement for as small a time interval as practicable (usually 3 or 6 hours between synoptic charts), using successive synoptic charts. The average of the velocity vectors at the beginning and the end of a given time interval would be taken as the true velocity and direction of the parcel over that time interval. Obviously, the smaller the time interval, the more accurate the trajectory. One of the charts may be a prognostic chart for computing future trajectories. Air trajectories are valuable in estimating the influence the earth has on air as it flows over varied earth surfaces.

Anticyclone. Also termed *high*, an atmospheric circulation having a sense of rotation about the local vertical opposite to that of the earth's rotation, i.e., clockwise in the northern hemisphere; counterclockwise in the southern hemisphere, undefined at the equator; a closed circulation, whose flow is within a closed streamline. With respect to the relative direction of its rotation, it is the opposite of a cyclone.

The barometric pressure within an anticyclone is high relative to its surroundings, and a pressure gradient exists from its center toward its periphery. A well-developed anticyclone is, essentially, an air mass, whose dimensions vary from a few hundred to several thousand miles (kilometers). It is, in general, a region of slowly settling air with a descent rate of from 300–1500 feet (90–460 meters) per day. Anticyclones are migratory in the region north of 30°–40° latitude, their path usually being to the east and south. Seasonal semipermanent anticyclones develop over both North America and Eurasia during winter. A belt of permanent anticyclones, with their centers usually over the oceans, lies between 10° and 40° latitude.

Anticyclones are generally accompanied by bright, clear weather believed to be the result of descending dry air at the anticyclone center; however, rain, drizzle, and cloudy skies may develop in the southwestern and western sectors of the air mass. Anticyclones moving from the north bring cold waves in winter, and cool, clear weather at other seasons. Those moving from the south bring mild weather in winter and hot, dry spells in summer.

Anticyclogenesis is any strengthening or development of anticyclonic circulation in the atmosphere. This applies to the development of anticyclonic circulation where, previously, it was nonexistent, as well as to intensification of existing anticyclonic flow.

Convention. Atmospheric motions that are predominantly vertical, resulting in vertical transport and mixing of atmospheric properties. **Autoconvection** is the phenomenon of the spontaneous initiation of convection in an atmospheric layer in which the lapse rate is equal to or greater than the autoconvective lapse rate. The presence of viscosity, turbulence, and radiative heat transfer usually prevents the occurrence of autoconvection until the lapse rate is greater than the theoretical autoconvective lapse rate of approximately $+3.4 \times 10^{-4}$°C per centimeter.

Cyclone. An atmospheric cyclonic circulation, a closed circulation. A cyclone's direction of rotation (counterclockwise in the northern hemisphere; clockwise in the southern hemisphere) is opposite to that of an anticyclone. While modern meteorology restricts the use of the term cyclone to the so-called cyclonic-scale circulations (with wavelengths of 1000 to 2500 kilometers), it is popularly applied to the more-or-less violent small-scale circulations such as tornadoes, waterspouts, dust devils, etc. (which may, in fact, exhibit anticyclonic rotation), and even, very loosely, to any strong wind. This term was first used very generally as the generic term for all circular or highly curved wind systems. Because cyclonic circulation and relatively low atmospheric pressure usually coexist, the terms cyclone and low are used interchangeably. Also, because cyclones nearly always are accompanied by inclement, and often destructive, weather, they are frequently referred to simply as storms.

Equatorial Vortex. A closed cyclonic circulation within the equatorial trough. It develops from an equatorial wave.

Ferrel Law. When a mass of air starts to move over the earth's surface, it is deflected to the right in the northern hemisphere, and to the left in the southern hemisphere, and tends to move in a circle whose radius depends upon its velocity and its distance from the equator.

Geopotential Height. The height of a given point in the atmosphere in units proportional to the potential energy of unit mass (geopotential) at this height, relative to sea level. The relation, in the centimeter-gram-second (c.g.s.) system, between the geopotential height Z and the geometric height z is

$$Z = \frac{1}{980}\int_{0}^{z} g \, dz$$

where g is the acceleration of gravity, so that the two heights are numerically interchangeable for most meteorological purposes. Also, one geopotential meter is equal to 0.98 dynamic meter. At the present time, the geopotential height unit is used for all aerological reports, by convention of the World Meteorological Organization. See also **Geopotential.**

Geopotential Surface. Also called equigeopotential surface or level surface, this is a surface of constant geopotential, i.e., a surface along which a parcel of air can move without undergoing any changes in its potential energy. Geopotential surfaces also coincide with surfaces of constant geometric height. Because of the poleward increase of the acceleration of gravity along a constant geometric-height surface, a given geopotential surface has a smaller geometric height over the poles than over the equator.

Gradient Flow. Horizontal frictionless flow in which isobars and streamlines coincide; or equivalently, in which the tangential acceleration is everywhere zero. Important special cases of gradient flow, in which two of the normal forces predominate over the third, are: (1) *Cyclostrophic flow*, in which the centripetal acceleration exactly balances the horizontal pressure force; (2) *Geostrophic flow*, where the Coriolis force exactly balances the horizontal pressure force; (3) *Inertial flow*, which is flow in the absence of external forces; in meteorology, frictionless flow in a geopotential surface in which there is no pressure gradient, so that centripetal and Coriolis accelerations must be equal and opposite.

Inertial Force. A term used specifically in meteorology to designate a force in a given coordinate system arising from the inertia of a parcel moving with respect to another coordinate system. For example, the Coriolis acceleration on a parcel moving with respect to a coordinate system fixed in space becomes an inertial force, the Coriolis force, in a coordinate system rotating with the earth.

Rossby Number. The nondimensional ratio of the inertial force to the Coriolis force for a given flow of a rotating fluid. It may be given as

$$\mathrm{Ro} = \frac{U}{fL}$$

where U is a characteristic velocity, f the Coriolis parameter (or, if the system is cylindrical rather than spherical, twice the system's rotation rate), and L a characteristic length.

The *thermal Rossby number* is the nondimensional ratio of the inertial force due to the thermal wind and the Coriolis force in the flow of a fluid that is heated from below.

$$\mathrm{Ro}_T = \frac{U_T}{fL}$$

where f is the Coriolis parameter, L a characteristic length, and U_T a characteristic thermal wind.

Solenoid. In meteorology, a tube formed in space by the intersection of unit-interval isotimic surfaces of two scalar quantities. Solenoids formed by the intersection of surfaces of equal pressure and density are frequently referred to in meteorology. A barotropic atmosphere implies the absence of solenoids of this type, since surfaces of equal pressure and density coincide.

Tangential Acceleration. The component of the acceleration directed along the velocity vector (streamline), with magnitude equal to the rate of change of speed of the parcel dV/dt, where V is the speed. In horizontal, frictionless atmospheric flow, the tangential acceleration is balanced by the tangential pressure force

$$\frac{dV}{dt} = -\alpha \frac{\partial p}{\partial s}$$

where α is the specific volume, p the pressure, and s a coordinate along the streamline. Thus, flow without tangential acceleration is along the isobars, and the wind is the gradient wind.

Circulation Theorem

If the atmosphere is baroclinic, that is, if the surfaces of equal pressure and equal density intersect at any angle whatsoever, there is a tendency for a circulation to develop in such a manner that the atmosphere will become barotropic, that is, the surfaces of equal pressure and equal density will coincide. The atmosphere is normally baroclinic. Sea and land breezes, mountain and valley breezes are results of well-defined baroclinic states. Direction of circulation is always such that cold air flows toward warm air at the base of the circulation pattern, and warm flows toward cold at the top of the pattern. Air sinks in the cold air region and rises in the warm air region. Thus, the sea breeze blows along the surface of the earth from cold water to heated land, rises, then returns seaward, and sinks.

It is possible to compute the magnitude of the circulation from a given baroclinic state if temperature and pressure are known. Suppose a vertical plane were erected perpendicular to a shore, extending from the cold water to heated land. Lines of equal pressure and temperature drawn in this plane will produce a field of approximate parallelograms, which are known as solenoids. See Fig. 6. A tendency for circulation exists about the perimeter of each solenoid, but, in the field as a whole, this tendency is nullified in adjacent solenoids. There is no nullification along the border solenoids of the whole field, and it is here that a circulation springs up. The number of solenoids in the field is a measure of the expected strength of the resulting circulation.

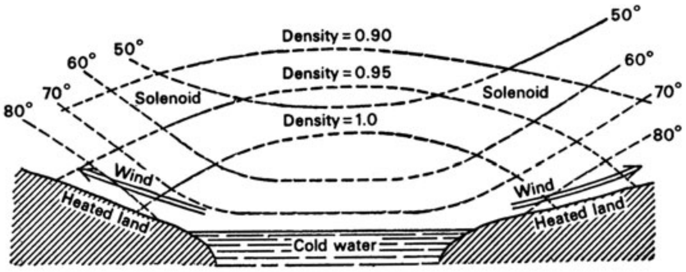


Fig. 6. Circulation theorem and solenoids.

Friction Between Wind and the Earth

Even though winds may blow with considerable velocity above the earth's surface, air cannot move rapidly in shallow layers just above the earth where air is caught in the irregularities of the earth's surface. This slowing of a wind at the earth's surface is the result of friction between the moving air and the earth. The effect of friction extends to about 500 meters ($>$ 1500 feet). The first effect of reduced velocity is a reduction in the Coriolis force and, therefore, a pressure gradient that is not balanced. Because the pressure gradient force is not balanced. air flows across isobars from high to low pressure. Thus, the unbalancing of the pressure gradient force causes air to converge toward low pressure and diverge from high pressure.

Over land, the average deflection very near the earth is up to 30° and over water it is about 20°. Reduction of wind speed depends upon the nature of the surface. Above the friction layer, the wind speed may be up to twice that at the surface.

Steady-State Wind Equation. In this equation, friction is ignored and the isallobaric and divergence-convergence factors do not come into play.

$$\frac{V^2}{r} + 2\Omega v \sin\phi = \alpha \frac{\partial p}{\partial r}$$

where v = wind speed
Ω = angular velocity of the earth
r = radius of curvature of the flow
α = specific volume, and is $1/\rho$, the density
$\frac{\partial p}{\partial r}$ = the pressure gradient along r
ϕ = the latitude

The wind speed in this relationship is given by:

$$v = \frac{1}{\rho}\frac{\partial p}{\partial r}\frac{1}{2\Omega \sin\phi}$$

where $\alpha = 1/\rho$ and ρ = density.

For straight-line flow,

$$v^2/r = 0,$$

and

$$v = -\Omega r \sin\phi \left[1 - \sqrt{1 - \frac{1}{r\Omega^2 \sin^2\phi}\frac{1}{\rho}\frac{\partial p}{\partial r}}\right]$$

For curved flow about an area of low pressure,

$$v = -\Omega r \sin\phi \left[1 \pm \sqrt{1 + \frac{1}{r\Omega^2 \sin^2\phi}\left(\frac{1}{\rho}\right)\left(\frac{\partial p}{\partial r}\right)}\right]$$

which is counterclockwise in the northern hemisphere and clockwise in the southern hemisphere where ϕ is negative.

For curved flow about an area of high pressure,

$$v = -\Omega r \sin\phi \left[\sqrt{1 + \frac{1}{r\Omega^2 \sin^2\phi}\frac{1}{\rho}\frac{\partial p}{\partial r}} - 1\right]$$

which is clockwise in the northern hemisphere and counterclockwise in the southern hemisphere.

Near the equator, the Coriolis force is approximately zero, and

$$v = \sqrt{\frac{r}{\rho}\frac{\partial p}{\partial r}}$$

Atmospheric Convergence and Divergence

If an imaginary box is erected in the atmosphere near the earth's surface, in such a manner that its base, top, and sides are parallel to the air flow (i.e., the winds), it is possible to illustrate the effects of convergence and divergence. See Fig. 7. When air flows uniformly through this box, there will be no accumulation or diminution of air inside the box. If, for any reason, however, more air flows into one end of the box than flows out the other, there is an accumulation of air, which must seek an outlet. Because the pressure is less at the top of the box than at the bottom, this accumulated air flows upward out of the box. If we

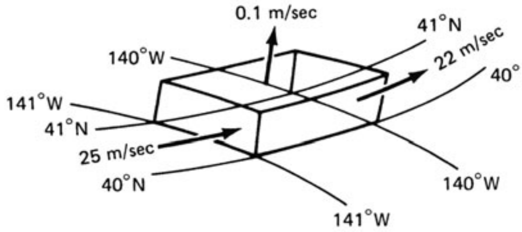


Fig. 7. Principle of convergence and divergence.

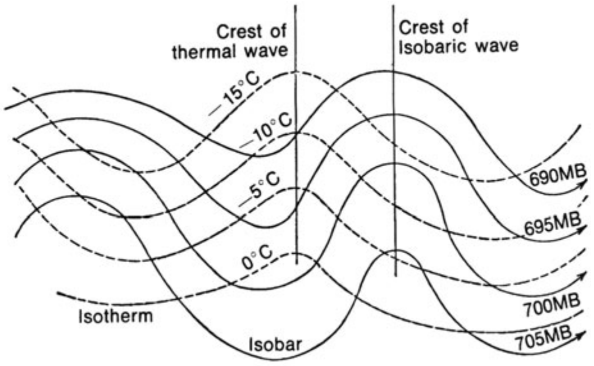


Fig. 8. Long waves in the westerlies.

followed a cube-shaped unit mass of air through this flow, it would become distorted into a rectangular prism elongated vertically. This process is called *convergence*, and results in a field of rising air. Converging, and therefore rising air can, if the process endures over a sufficient period of time, produce clouds and precipitation. It also tends to destabilize the air. If, in the same box, less air flows into one end than flows out the other, there is an air diminution, and space is available at the top of the box for more air. One unit cube of air will be flattened into a rectangular prism elongated horizontally. This process is called *divergence*, and results in a field of sinking or subsiding air. Divergence, therefore, and subsiding air tend to stabilize the air. Clouds and turbulence diminish in regions of divergence and subsidence.

Waves in the Atmosphere

In meteorology, any pattern with some roughly identifiable periodicity in time and/or space applies to atmospheric waves in the horizontal flow pattern, i.e., in the wind field, there are wavelike disturbances, such as equatorial, easterly, frontal, Rossby, long, short, cyclone, and barotropic waves. The study of water-surface waves has bred its own special terminology, such as deep-water, shallow-water, wind, hurricane, and tidal waves. In popular terminology, a surge or influx, is often referred to as a wave, i.e., "heat wave," or "cold wave."

Barometric waves are any waves in the atmospheric pressure field, usually reserved for short-period variations not associated with cyclonic-scale motions or with atmospheric tides.

Barotropic waves occur in a two-dimensional, nondivergent flow, the driving mechanism lying in the variation of vorticity of the basic current and/or in the variation of the vorticity of the earth about the local vertical.

Cyclone waves are (1) disturbances in the lower troposphere, of wavelengths from 1000 to 2500 kilometers, recognized on synoptic charts as migratory high- and low-pressure systems, and identified with the unstable perturbations connected with baroclinic and shearing instability; and (2) frontal waves at the crests of which are centers of cyclonic circulation; therefore, the frontal waves of wave cyclones.

Easterly waves are migratory wavelike disturbances of the tropical easterlies, which move within the broad easterly current from east to west, generally more slowly than the currents in which they are imbedded. Although best described in terms of wavelike characteristics in the wind field, they also consist of weak troughs of low pressure. Easterly waves do not extend across the equatorial trough.

Equatorial waves are wavelike disturbances of the equatorial easterlies that extend across the equatorial trough.

Frontal waves are horizontal wavelike deformations of fronts in the lower levels, commonly associated with a maximum of cyclonic circulation in the adjacent flow. They may develop into wave cyclones.

Gravity wave disturbances are those in which buoyancy (or reduced gravity) acts as the restoring force on parcels displaced from hydrostatic equilibrium. There is a direct oscillator conversion between potential and kinetic energy in the wave motion.

Long Waves in the Prevailing Westerlies. There develop in the westerlies, particularly during the cold months, certain perturbations which cause the westerlies to blow alternately northward and southward in a sinusoidal wave pattern, but always with a component of velocity directed from west to east. See Fig. 8. Ridges are associated with anticyclones at ground and near-ground levels, whereas troughs are associated with cyclones. There is, therefore, a definite relation between the sinusoidal perturbations of the westerlies and large-scale surface weather phenomena. Progression eastward, or retrogression westward, of the crests and troughs is usually about the same as surface anticyclones and cyclones. Sinusoidal perturbations can, therefore, be used for prognostic purposes computing their velocities. These velocities are given by the formula:

$$C = U - \frac{bL^2}{4\pi^2}$$

where C = the velocity of the wave
U = the west to east component of air motion
L = the wavelength of the wave
b = the rate of change in the Coriolis parameter northward

Wave velocities may be positive, zero, or negative; positive indicating easterly movement, negative, westerly movement.

Isotherms aloft also assume partial or complete sinusoidal form under some conditions. The following relations between amplitudes of the thermal and the streamline wave apply:

$$C = U\left(1 - \frac{A_s}{A_t}\right)$$

where A_s = streamline amplitude
A_t = thermal amplitude

From this, the following conclusions are possible:

1. If the amplitude of the thermal wave is greater than the streamline wave and the two are in phase, cyclones and anticyclones will have a slow eastward component of movement.
2. If the thermal wave is 180° out of phase with the streamline wave, cyclones and anticyclones move eastward rather rapidly. The smaller the amplitude of the thermal wave in relation to the streamline wave, the more rapid will be the movement of surface systems.
3. If the amplitude of the thermal wave is less than the streamline wave and the two are in phase, cyclones and anticyclones will move with a slow westward component (retrograde motion).
4. If the amplitudes of the thermal and streamline waves are the same and the waves are in phase, the surface systems will have no east-west component of movement.

The first two of these conditions are common: the latter two occur but are not usual.

Mountain Waves in the Atmosphere. These are internal waves located between the earth's surface and the base of the stratosphere. They are created when winds blow across mountainous terrain with a speed of 50 knots or more at the height of the tallest terrain and simultaneously the vertical structure of the atmosphere is favorable for wave development and growth. In North America, the most common locations for development of mountain waves are along the Rocky Mountains, over the Sierra Nevadas, and over the Pacific Coastal Range.

Within the atmosphere when mountain waves are present, the air surges upward and downward; there are vortices and the air movement is chaotic. Evidence of mountain waves is found in the presence of standing lenticular clouds, occasionally a patch of dust kicked up by a gust and, in some select sites, strong gusty winds in mountain passes and canyons.

The most serious impact of mountain waves in the atmosphere is on aircraft that happen to be in the airspace where mountain waves are present. Strong updrafts and downdrafts and moderate-to-severe turbulence are the common experiences.

Conditions under which mountain waves are generated are relatively well known and understood. The wind direction should lie between 90° and not less than 45° to the crest of the terrain. The wind speed should be 50 or more knots at the top of the terrain and should increase with altitude. There should be a stable layer of air or an inversion approximately 10,000 feet (3048 meters) above the level of high terrain. A low tropopause is also favorable.

There are many small local areas where mountain waves occur when the structure of the atmosphere and the winds are favorable. Most prominent locations in the contiguous states of the United States include:

1. Northeastern New Mexico to southeastern Wyoming, including all of central eastern Colorado. Winds are westerly.
2. Nevada-California border and over the western part of Nevada. Winds are west southwesterly.
3. Central northeast Wyoming through central Montana. Winds are southwesterly.

Periodic Changes in the Atmosphere

The atmosphere undergoes periodic recurring changes associated primarily with the relative position of the sun, but to a lesser amount on other factors. The most prominent and obvious periodic changes are the diurnal changes. Less dramatic, but nonetheless obvious are the seasonal changes. A periodic, but recurring changes are associated with the pattern of circulation in both hemispheres, and may also be dependent in part on the changes in ocean surface temperatures. There are also tides in the atmosphere which can be detected with very sensitive equipment.

Diurnal Changes. Changes completed within and recurring every 24 hours. The diurnal variability of nearly all meteorological elements is one of the most striking and consistent features of the study of weather. The diurnal variations of important elements at the earth's surface can be summarized as follows:

1. Atmospheric pressure varies diurnally or semidiurnally according to the effects of atmospheric tides. Surface pressure undergoes two definite periods of increase and two of decrease. Mean maximum pressure occurs approximately at ten o'clock local time, in the morning and evening; and mean minimum pressure occurs at four o'clock in the afternoon and morning. In the tropics, this surge and ebb of pressure is very pronounced and highly rhythmic.
2. Temperature tends to reach its maximum about 2–3 hours after local noon, and its minimum at sunrise. Over water, there is a minimum diurnal change as small as a fraction of a degree, and over sandy and rocky desert a maximum that sometimes amounts to 100° or more.
3. Relative humidity tends to become maximum about sunrise, and minimum in the afternoon.
4. Cloudiness and precipitation over a land surface increase by day and decrease at night; over water, the reverse is true but to a lesser extent. Over land, cumulus-type clouds tend to be maximum during afternoons and minimum at night.
5. Fogs tend to be maximum at and shortly after sunrise, and minimum in the afternoon.
6. Evaporation is markedly greater by day; and condensation is much greater at night.
7. Wind generally increases and veers by day, and decreases and backs by night; rough flying air tends to be maximum in midafternoon, and minimum at night.
8. Onshore winds tend to build up during the later morning and afternoon—then die out again in the evening and at night. Offshore winds sometimes set in during nighttime hours. Valley breezes tend to develop—blowing toward higher terrain by day and then reverse to blow downhill by night.

Seasonal Changes. These changes complete their cycle during a year and recur with some variations in intensity during the same months each year. The most prominent and obvious seasonal changes are snow in the winter and thunderstorms in the summer. Some of the more important seasonal changes include:

1. Snow occurs only in winter in most inhabited areas north of 30°N or south of 30°S.
2. The average temperature, mean maximum, and mean minimum temperatures increase to a high value during the second and third months of summer and decrease to a low value during the latter part of winter.
3. Thunderstorms increase in numbers to a maximum occurrence in summer.
4. Fog and low clouds are at a maximum during winter in temperate zone areas.
5. Tornadoes in those areas where they tend to occur are at a maximum in number and intensity during spring months.
6. Severe tropical storms tend to reach a maximum in number during autumn months.
7. Large-scale onshore monsoon circulations are predominantly a summer phenomenon whereas the offshore circulation occurs in winter.
8. Prolonged periods of rainfall (wet season) in the tropics and subtropics are associated with summer; dry periods with winter.

Aperiodic Changes. These appear to be linked to the pattern of circulation of the westerly flow in the atmosphere. There are a number of stable configurations of the west-to-east flow patterns in the atmosphere. Each one of the patterns consists of a discrete number of waves and each wave has a crest and trough. The crests and troughs tend to remain stationary for varying periods of time from weeks to months.

Air movement in the airspace between the trough and the crest of a wave (looking eastward) has a component from the equator toward the poles which tends to carry moisture. Cloud and precipitation-bearing storms predominantly are found in this zone. The airspace between the crest and the trough has an air movement component from the polar region toward the equator which has an associated suppression of cloudiness and precipitation.

A particular stable configuration may remain essentially unchanged for a substantial period of time, causing drought or cold or heat in one area and opposite conditions in another area. When the pattern breaks down and a different stable configuration is established, the newly established pattern will shift associated conditions to different areas. Shifts from one stable configuration to another are aperiodic (irregular).

Atmospheric Tide

Also known as atmospheric oscillation, the atmospheric tide is an atmospheric motion of the scale of the earth, in which vertical accelerations are neglected (but compressibility is taken into account). Both the sun and moon produce atmospheric tides, which may be thermal or gravitational. A *gravitational tide* is due to gravitational attraction of the sun or moon; the semidiurnal solar atmospheric tide is partly gravitational; the semidiurnal lunar atmospheric tide is fully gravitational. A *thermal tide*, so-called in analogy to the conventional gravitational tide, is a variation in atmospheric pressure due to the diurnal differential heating of the atmosphere by the sun.

The amplitude of the *lunar atmospheric tide* is so small (about 0.06 millibar in the tropics and 0.02 millibar in middle latitudes) that it is detected only by careful statistical analysis of a long record; the only detectable components are the 12-lunar-hour or semidiurnal, as in the oceanic tides, and two others of very nearly the same period. The 12-hour harmonic component of the *solar atmospheric tide* is both gravitational and thermal in origin, and has, by many times, the greatest amplitude of any atmospheric tidal component (about 1.5 millibars at the equator and 0.5 millibar in middle-latitudes); 6- and 8-hour tides of small amplitude have been observed, as well as the 24-hour component, which is a thermal tide with great local variability. The fact that the 12-hour component of the solar atmospheric tide is greater than the corresponding lunar atmospheric tide is ascribed usually to a resonance in the atmosphere with a free period very close to the tidal period.

Peter E. Kraght, Certified Consulting Meteorologist,
Mabank, Texas.

ATMOSPHERE-OCEAN INTERFACE. Almost 71% of the earth's atmosphere is in contact with oceanic surfaces. The ocean-air interface, therefore, plays a dominant role in determining the water content and

TABLE 1. PERCENT OF LAND AND OCEAN FOR SPECIFIED LATITUDE BELTS.

Latitude Belt	Northern Hemisphere					Southern Hemisphere				
	Percent		Average Ocean Surface Temperature, °C			Percent		Average Ocean Surface Temperature, °C		
	Water	Land	Pacific	Atlantic	Indian	Water	Land	Pacific	Atlantic	Indian
80–90	92.6	7.4					100			
70–80	71.3	28.7				24.6	75.4			
60–70	29.9	70.1		5.6		89.6	10.4	−1.3	−1.3	−1.5
50–60	42.8	57.2	5.7	8.7		99.2	0.8	5.0	1.8	1.6
40–50	47.5	52.5	10.0	13.2		96.9	3.1	11.2	8.7	8.7
30–40	57.2	42.8	18.6	20.4		88.8	10.2	17.0	16.9	17.0
20–30	62.4	37.6	23.4	24.2	26.1	76.9	23.1	21.5	21.2	22.5
10–20	73.6	26.4	26.4	25.8	27.2	78.0	22.0	25.1	23.2	25.8
00–10	77.2	22.8	27.2	26.7	27.9	76.4	23.6	26.0	25.2	27.4
00–90	60.7	39.3				80.9	19.1			

temperature of the lower levels of the atmosphere and perhaps of the total atmosphere. The percentage of land and ocean for specified latitude belts and the temperatures in the main oceans in each zone are given in Table 1.

A broad zone lying roughly between 30°N and 30°S is approximately 75% water, with an annual average temperature of near 25°C (77°F) and nowhere below 20°C (68°F). Air in contact with this broad expanse of warm water acquires the properties of the water surface and is permeated by water vapor to considerable depths. This zone is the source region of moist, warm, unstable tropical air masses that move toward the poles as part of the general circulation of large-scale cyclones and anticyclones.

Source of Water Vapor. The air-ocean interface is the primary source of water for the atmosphere and particularly between the latitudes of 40°N and 40°S where the temperature averages 20°C (68°F) or higher. The average inches of liquid water evaporated each year into each vertical tube of air one-inch square in contact with the ocean surface is given in Table 2. This water vapor is carried aloft and transported laterally within the atmosphere. Secondary sources of water vapor entering the atmosphere include transpiration from plants, evaporation from moist soil and rock, lakes, and rivers.

TABLE 2. LIQUID WATER EVAPORATED BY OCEANS.
(Average Inches of Liquid Water Evaporated Each Year Into Each Vertical Tube of Air One-Inch Square in Contact with the Ocean Surface)

Latitude	Northern Hemisphere	Southern Hemisphere
40°	37 inches	32 inches
35°	42	39
30°	47	43
25°	51	49
20°	52	53
15°	51	53
10°	50	51
05°	43	49
00°	47	47

Thermal Stabilization of the Atmosphere. As provided by the oceans, this is caused primarily by the large-scale uniformity of ocean temperature at the air-ocean interface. From day to day there is no appreciable change in ocean surface temperature except locally in immediate offshore waters. There is a slow seasonal change. Air in contact with the thermally stable ocean surfaces also tends to become thermally stable. Annual seasonal change in average ocean surface temperature for the Pacific, Atlantic, and Indian Oceans is indicated in Table 3. Were it not for the stabilizing influence of the oceans, the annual temperature range would be much greater than it is, i.e., warmer in summer and colder in winter.

TABLE 3. ANNUAL SEASONAL CHANGE IN AVERAGE OCEAN SURFACE TEMPERATURES.

Latitude	Northern Hemisphere Ocean						Southern Hemisphere Ocean					
	Pacific		Atlantic		Indian		Pacific		Atlantic		Indian	
	°F	°C	°F	°C	°F	°C	°F	°C	°F	°C	°F	°C
60°			9	5			4	2.2	4	2.2	4	2.2
50°	13	7.2	11	6.1			7	3.9	5	2.7	5	2.7
40°	18	10	14	7.7			9	5	9	5	7	3.9
30°	13	7.2	13	7.2			7	3.9	11	6.1	11	6.1
20°	7	3.9	7	3.9			5	2.7	7	3.9	7	3.9
10°	4	2.2	2	1.1	5	2.7	4	2.2	6	3.4	4	2.2
00°	4	2.2	4	2.2	2	1.1	4	2.2	4	2.2	2	1.1

Salt Condensation Nuclei. These nuclei required for cloud formation and precipitation originate through the air-ocean interface. Ocean spray from breaking waves creates tremendous numbers of small salt water droplets which are carried up into the atmosphere to evaporate and to leave a very small residue of sea salt. Wherever in the atmosphere the relative humidity approaches saturation values, these nuclei become the core for haze, small cloud droplets, and rain. Constituent chemical elements found in ocean water are also found in the same proportionate ratios in microscopic salt nuclei and in rain water. Over land areas, other nuclei from various sources tend to outnumber the salt nuclei.

Ocean Waves. These are predominately wind-generated. When the wind speed is near calm, the ocean surface is only rippled. As the wind speed increases, air at the ocean-air interface drags the water forward, causing waves to develop. The *fetch* is the stretch or distance over which winds can act upon the ocean surface. Fetch is limited by the distance of open water and by the distance that winds blow in one direction at a sufficient speed. Only rarely does wind direction stay in one orientation for long distances. Likewise, sustained high wind speeds only rarely extend over long distances.

Wave height is empirically related to wind speed by:

$$H = 0.025V^2$$

where H is in feet and V is in knots. The maximum wave height is related to the fetch by:

$$H_{\max} = 1.5\sqrt{F}$$

where $H_{\max}$ is in feet and F is in nautical miles. See Table 4.

TABLE 4. WAVE HEIGHT VERSUS WIND SPEED AND MINIMUM FETCH.

Wave Height (Feet)	Wind Speed (Knots)	Minimum Fetch Required To Attain Height (Nautical Miles)
3	10	3
10	20	48
23	30	240
42	40	760
65	50	1860

Fetches greater than 1000 nautical miles are not common and those as much as 2000 miles probably do not occur. The tallest waves observed are on the order of 60 to 70 feet (18 to 21 meters). Wave heights of 50 feet (15 meters) can be expected in winds of more than 45 knots provided the fetch is sufficient. Very large waves can develop in strong winds within 12 hours when the fetch is sufficiently long. In contrast, strong winds cannot create huge waves on a short fetch.

Swells. These are more or less uniformly spaced, rounded waves that were generated as wind waves, but that have traveled well beyond the ocean areas where they were developed. The orientation of swells changes only slowly, if at all, with time. Therefore, swells can be used to indicate the location of the storm that generated them. Observations of swells is usually part of a weather observation from ships at sea. The height, orientation, and speed of swells observed at a number of points in the open ocean provide meteorologists with useful information in pinpointing storm centers.

The ocean surface most often displays a chaotic mixture of swells and locally generated wind-driven waves. There may be as many as a half-dozen intermingling swells and waves in one spot, causing the sea surface to rise and fall in a most irregular manner.

Superwaves. In 1983, researchers J. G. and G. W. Moore (U.S. Geological Survey) investigated how certain boulders of limestone found on the Hawaiian island of Lanai reached such unusual heights. Limestone-bearing gravel, the newly named *Hulopoe Gravel*, blankets the coastal slopes of Lanai. The deposit, which reaches a maximum altitude of 326 meters (1070 feet), formerly was believed to have been deposited along several different ancient marine strand-lines, but dated submerged coral reefs and tide-gage measurements indicate that the southeastern Hawaiian Islands sink so fast that former worldwide high stands of the sea now lie beneath local sea level. Evidence indicates that the Hulopoe Gravel and similar deposits on nearby islands were deposited during the Pleistocene by a giant wave generated by a submarine landslide on a sea scarp south of Lanai.

Because of the great run-up of the wave, it was probably not a seismic sea wave caused by a subsea earthquake. Either the impact of a meteorite on the sea surface or a shallow submarine volcanic explosion could have generated the Hulopoe wave. The researchers believe, however, that a more likely explanation is a rapid downslope movement of a subsea landslide on the Hawaiian Ridge, which is among the steepest and highest landforms on earth. The occurrence of several major subsea landslides of various ages, possibly triggered by local earthquakes, indicates that the Hawaiian Ridge is a site of repeated slope failure. A landslide in a confined fjord in Alaska in 1958 produced a run-up of 524 meters, the highest on record. The researchers infer that that rapid movement of a submarine slide near Lanai displaced seawater forming a wave that rushed up onto the island, carrying with it rock and reef debris from the nearshore shelf and beach.

The El Niño 1982–83 Event

For well over a century, Ecuadorian and Peruvian fishermen have referred to the annual appearance of warm water in the Pacific off their shores at Christmas time as *El Niño* (Spanish for the Christ child). In meteorological terms, El Niño (EN) is an anomalous warming of surface water in the equatorial Pacific, mainly off the coast of Peru. Warming of these waters is a normal event, but periodically and difficult to forecast, the EN in some years is of much greater intensity, covers a larger area of water, and is prolonged. Instead of lasting for just a few months, the warm waters may persist for a year or longer, as in the case of a number of past instances: 1953, 1957–58, 1965, 1972–73, and 1982–83. During the latter event, the sea surface temperature off Peru rose by over 7 degrees Celsius (12.6°F). A temperature this high seriously affects the Peruvian anchovy fisheries. Traditionally, in a normal year, the catch will exceed 12 million tons. In the 1982–83 EN event, the annual catch dropped to less than one-half million ton. Michael Glantz of the National Center for Atmospheric Research (NCAR) takes exception that all of the consequences suffered by the Peruvian fishing industry be charged against the EN event, but rather reductions in the catch coincidentally also resulted from technological advances, political changes in the national government, and a lack of government-agency supervision of the fishing industry. Be that as it may, the EN was a major factor in the loss.

EN has been linked with a variety of atmospheric anomalies. Some are local—heavy rains in usually arid regions along the Pacific Coast of South America. EN also must be regarded as part of an interrelated set of changes in atmospheric and ocean conditions over much of the Southern Hemisphere, often referred to as the EN-Southern Oscillation, or simply *Southern Oscillation.* The latter may be defined as a massive seesawing of atmospheric pressure between the southeastern and the western tropical Pacific. Southern Oscillation involves a periodic weakening or disappearance of the trade winds, which triggers a complex chain of atmosphere-ocean interactions.

As observed by Michael Wallace (University of Washington), "Most of us feel that (EN) is a coupled phenomenon. The atmosphere itself doesn't have enough of an attention span to know what happened a couple of months ago. The ocean can serve as a memory. It can remember what happened a season or a year ago. But the atmosphere, unlike the ocean, has the large-scale systems that make this phenomenon global."

The apparent coupling between EN and global weather anomalies seemed very convincing in 1982–83 when EN was exceptionally intense. During this period, unusually severe Pacific storms struck the California, Oregon, and Washington coasts. These storms dumped heavy snow on western U.S. mountains, and spring floods followed. Extreme droughts hit many parts of the world, including the western Pacific and Mexico, and torrential rains and flooding drenched parts of South America and the southern United States. A number of scientists have attributed these extreme events to the extraordinary EN.

In an ordinary year, many storm systems form or intensify near the east coast of Asia and move across the Pacific Ocean. Eventually, the storms cross the western United States and continue eastward. However, in some EN years, the Pacific storm track veers northward toward Alaska, altering the usual paths of these winter storms. This happened in the winter of 1976–77, when the western United States had an unusually warm, dry winter, while severe cold and snow swept down over the eastern part of the country as far as Florida.

In 1983, John Geisler (University of Utah) and his colleagues used what is known as the community climate model (CCM) to simulate EN events of three different intensities. Although the model produced the most significant features of northern winter anomalies that accompany intense ENs, it is not clear why one EN winter can differ so much from another. There are indications that the difference may be related to the geographical location of the warm surface water.

Although this experiment may sound simple, it required considerable resources and effort. At the NCAR, its high-speed CRAY-1 computer made it possible. The experiment was nearly five times larger than any prior ones and required more than 100 hours of computer time.

Numerous meteorological and oceanographic scientists throughout the world found the timing right for tackling the challenge of understanding the 1982–83 EN. Much was learned, but scientists are not ready to acknowledge that they can forecast with any degree of reliability when the next EN event may return, or what its consequences may be. Some interesting theories have been proposed, but remain untested. Geoffrey Vallis (Scripps Institution of Oceanography) observed in 1986 that most of the principal qualitative features of the EN-Southern Oscillation phenomenon can be explained by a simple but physically motivated theory. These features are the occurrence of sea-surface warmings in the eastern equatorial Pacific and the associated trade wind reversal; the aperiodicity of these events; the preferred onset time with respect to the seasonal cycle; and the much weaker events in the Atlantic and Indian Oceans. The theory, in its simplest form, is a conceptual model for the interaction of just three variables, namely, (1) near-surface temperatures in the east, and (2) west equatorial ocean, and (3) a wind-driven current advecting the temperature field. For a large range of parameters, the model is naturally chaotic and aperiodically produces EN-like events. For a smaller basin, representing a smaller ocean, the events are proportionately less intense. Vallis summarizes by observing that although the model has many limitations, one being that it cannot describe spatial variations in any detail, it does explain many of the qualitative features of EN. It transparently demonstrates the underlying dynamics and thereby the possibility of a purely internal mechanism for the phenomenon. It shows that external triggering or stochastic forcing is not necessarily essential, although such effects may have a role in the real system.

Cane and Zebiak (Lamont-Doherty Geological Observatory) created a sophisticated model wherein the atmosphere and ocean are coupled to each other, allowing changes in one to affect the other. However, compared with a number of other models, the Cane-Zebiak model is relatively simple, but the designers claim that all essential ingredients are present. One of these is the potential for a feedback between the temperature gradient of the equatorial Pacific, its eastern end normally being colder than its western end, and the east-to-west winds normally blowing along the equator. Inasmuch as the temperature gradient drives these winds that normally keep the warm water at bay in the west, a warming in the east would weaken the gradient and thus the winds. This would lead to further warming. Positive feedback as described has been a part of most EN models since that created by Bjerknes in the 1960s, which related EN and its associated atmospheric phenomenon, the Southern Oscillation. Cane and Zebiak did include one other essential element, i.e., the need for the heat content of the upper tropical Pacific to be higher than normal. Without that precondition, the researchers believe that the feedback needed could not occur. An EN event predicted by the model for 1986–87 did not happen.

Biological Consequences of El Niño 1982–83. The EN event not only provided an excellent opportunity for atmosphere and ocean scientists to exercise new sophisticated and computerized equipment in the laboratory, but also to observe the EN by satellite. As reported by Fiedler (National Oceanic and Atmospheric Administration, La Jolla, California), satellite infrared temperature images illustrated several effects of the EN event. Warm sea-surface temperatures, with the greatest anomalies near the coast, were observed, as were weakened coastal upwelling, and changes in surface circulation patterns. Phytoplankton pigment images from the *Coastal Zone Color Scanner* indicated reduced productivity, apparently related to the weakened coastal upwelling. The satellite images provided direct evidence of mesoscale changes associated with the oceanwide EN event.

In addition to the previously mentioned EN effects of severely lowering the Peruvian anchovy catch, strong winter storms in southern California, attributed by many scientists to the EN event, destroyed most of the canopy of the giant kelp *Macrocystis pyrifera.*

As reported by Michael Glantz in a 1984 paper, some scientists referred to the EN 1982–83 event as the most potent in about a century, because of its alleged linkages to the devastating impacts on the economies of countries that border the Pacific Ocean in the Southern Hemisphere—droughts in Australia, Indonesia, Peru, and Hawaii; an increase in the number of destructive tropical typhoons in the southern Pacific region; the mysterious disappearance and subsequent reappearance of seabirds on Christmas Island; the destruction of Pacific coral reefs; and the decimation of fish stocks that normally inhabit the coastal waters of Peru and Ecuador, (in addition to anchovy).

Just which of the aforementioned effects were directly or indirectly attributable to EN-1982–83 will remain controversial until improved computerized models are developed and tested once again at the occurrence of another intense EN in the indefinite future.

ATMOSPHERE (Planetary). See **See specific planets.**

ATMOSPHERIC INTERFERENCE (or Spherics). The interference of radio reception caused by natural electric disturbances in the atmosphere.

ATMOSPHERIC OPTICAL PHENOMENA. Because of varying conditions present in the atmosphere from time to time and notably the presence of ice crystals, dust particles, and other particulate matter, several interesting optical effects are the result.

Sky Color. The characteristic blue color of clear skies is due to preferential scattering of the short-wavelength components of visible sunlight by air molecules. Presence of foreign particles in the atmosphere alters the scattering processes in such a way as to reduce the blueness. Hence, spectral analysis of diffuse sky radiation provides useful information concerning the scattering particles. The study and measurement of the blueness of the sky is called *cyanometry.* Sometimes the Linke scale (or blue-sky scale) is used. The Linke scale is simply a set of eight cards of different standardized shades of blue. They are numbered 2 to 16, the odd numbers to be used by the observer when the sky color appears to lie between any of the given shades.

Halos and Coronas

A halo is any one of a large class of atmospheric phenomena appearing as colored or whitish rings and arcs about the sun or moon when seen through an ice crystal cloud, or in a sky filled with falling ice crystals. The halos exhibiting prismatic coloration are produced by refraction of light by the crystals and those exhibiting only whitish luminosity are produced by reflection from the crystal faces. The minute spicules of ice, in falling, take some definite attitude determined by their shape. Some are needlelike and assume a horizontal position; some are flat disks or stars and fall with their planes horizontal; while others, made up of both disks and rods, behave like a parachute. The sunlight is refracted by each type in a characteristic manner and dispersed into colors; it is also reflected from their external surfaces without dispersion.

Halos differ from coronas in that the former are produced by refraction and reflection due to ice crystals, whereas the latter are produced by diffraction and reflection due to water drops. A colored halo may often be distinguished from a corona in that it has the red nearest the sun or moon, whereas the corona has red in the exterior rings.

Halo Phenomena. With regards to type, orientation, motion, and solar elevation angle of ice crystals, a large variety of halos is possible, theoretically. Many varieties have been observed. Some halos theoretically predicted have not yet been reported; some that have been reported have not yet been theoretically explained, such as the Hevelian halo (described below). On rare occasions, an observer's sky will be filled with a display of four or five halo phenomena at one time, usually persisting for only a few minutes. Much supernatural lore was built up about such displays by the ancients.

By far the most common halo phenomenon is the *halo of 22°*, in the form of a prismatically colored circle of 22° angular radius around the sun or moon, exhibiting coloration from red on the inside to blue on the outside. It is produced by refraction of light that enters one prism face and leaves by the second prism face beyond, thus being refracted by an effective prism of 60° angle. In order to have a full 22° halo, the sky must be filled with hexagonal ice crystals falling with random orientations, a condition that apparently is frequently satisfied. This halo exhibits a distinct spectral pattern out to blue, due to the tipping of the crystals and consequent overlap of spectra in all but the red end. A reddish inner edge is usually all one can discern.

Closely allied to the 22° halo are the *parhelia* ("mock suns" or "sun dogs") and *paraselenae* ("mock moons"). The parhelia are two colored (reddish) luminous spots that appear at points 22° (or somewhat more) on both sides of the sun and at the same elevation as the sun. Their lunar counterpart is the weakly colored paraselenae, which are observed less frequently than the parhelia because of the moon's comparatively weak luminosity. These phenomena are produced by refraction in hexagonal crystals falling with principal axes vertical, the effective prism angle being 60°, as in the halo of 22°.

The parhelic circle ("mock sun ring") and the paraselenic circle ("mock moon ring") are halos consisting of a faint white circle passing through the sun or moon, respectively, and running parallel to the horizon for as much as 360° of azimuth. These circles are often seen in the sky along with parhelia or paraselenae, and are produced in the same manner. Parhelia and paraselenae occur at several positions along the parhelic circle other than the common 22° position, i.e., at 46°, 90°, 120°.

The *Hevelian halo* is the halo of 90°; it appears, only occasionally, as a faint, white halo on the sun or moon, and is a member of the class of halos reported but not yet fully explained.

A *sun pillar* (or a "light pillar") is a luminous streak of light, white or slightly reddened, extending above and below the sun, most frequently observed near sunrise or sunset. It may extend to about 20° above the sun, and generally ends in a point. The luminosity is thought to be produced simply as a result of reflection of sunlight from the tops and bottoms of tabular hexagonal ice crystals falling with principal axes vertical.

A *sun cross* is a rare halo phenomenon in which bands of white light intersect over the sun at right angles. It appears probable that most of such observed crosses appear merely as a result of the superposition of a parhelic circle and a sun pillar.

The *arcs of Lowitz* is a type of halo, rarely seen, in which the luminous arcs extend obliquely downward from the 22° parhelia on either side of the sun (or moon); it is concave towards the sun, with reddish inner edges; and is produced by refraction in hexagonal ice crystals that are oscillating as they descend.

The *circumhorizontal arc* is produced by refraction of light entering snow crystals, and consists of a colored arc, red on its upper margin, which extends for about 90° parallel to the horizon and lies about 46° below the sun. The *circumzenithal arc* is produced by refraction of light entering the tops of tabular ice crystals; it consists of a brightly colored arc about 90° in arc length, and is found about 46° above the sun, with its center at the zenith. It is typically very short-lived, but also very brilliant. In addition to these arcs, several types of halo arcs known generically as *tangent arcs* are occasionally formed as loci tangent to other halos, especially to the halo of 22°.

Corona. The corona consists of one or more rings located symmetrically about the sun or moon caused by diffraction of light passing through liquid water droplets. Coronas are of varied radii about the sun or moon, dependent upon the size of the water droplets. The radius of any corona is inversely related to the diameter of the water droplets causing it:

$$\sin \theta = \frac{(N + 0.22)L}{D}$$

where θ = angular radius of the corona ring
N = order of the corona ring (1st, 2nd, etc.)
L = wavelength of the light
D = diameter of the water droplets

The order of coloration in a corona ring is from blue on the inside to red on the outside (opposite to the coloration of halo rings).

Bishop rings are corona rings of faint reddish-brown seen in dust clouds.

Glory or Anticorona. A glory ring is observed on a cloud top of edge opposite to the position of the sun, i.e., the antisolar point. Glories are most frequently observed from aircraft flying above clouds. The shadow of the plane is in the center of the glory ring. See Fig. 1. These anticorona rings are complementary to the corona rings.

Bouguer's halo is a ring of faint white light usually about 39 feet (12 meters) in radius observed on some occasions outside the glory ring.

Rainbow

Looking into a "sheet" of water drops (rain, fog, spray) that is illuminated by strong white light from behind, an observer sees one, and sometimes two, concentric, spectrally colored rings, called a rainbow. If two are visible, the inner ring, called the "primary bow," is brighter and narrower than the outer ring or "secondary bow." In the primary bow, red is on the outside edge and violet on the inside edge; the order in the secondary bow is reversed. The colors are not so pure as in a spectrum because each wavelength extends over a wide radial range, the rainbow itself being made up of the fairly pronounced intensity maxima.

Fig. 1. Photograph of a glory ring made by a crew member of an American airliner. Note that center of glory ring is at exact position of cockpit in shadow of aircraft.

The colors of the rainbow are caused by the refractive dispersion of the spherical water drops. In Fig. 2 is shown the dispersion composing the primary and secondary bow. The diagram also explains why the order of colors is reversed, and shows that only the highest drops in the primary bow, and the lowest in the secondary bow, refract red light to the eye. The two internal reflections, with consequently greater loss of light, explain why the secondary bow is fainter. The center of the ring system is exactly opposite the source of light, so that natural rainbows are seen only when the sun is near the horizon, unless the observer is elevated high above the surrounding country and can look obliquely downward into the rain.

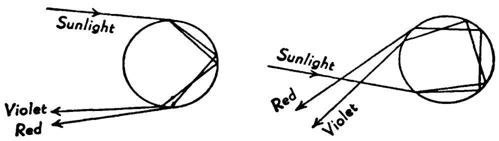


Fig. 2. Formation of primary bow (left) and secondary bow (right). Circles represent a raindrop.

A *fog bow* is a type of rainbow, faintly colored, seen on fog layers whose droplets are small.

Mirage

A mirage is a curious atmospheric phenomenon caused by the total reflection of light at a layer of rarefied air. The most familiar manifestation is observed in warm weather on paved highways. The air next to the pavement becomes heated and rarefied in comparison with that above it, so that, at a sufficient angle of incidence, objects beyond the area are mirrored as if by polished silver, giving the almost irresistible impression that one is looking at a layer of water. Travelers in hot desert regions are sometimes thus deceived. Much more rarely the phenomenon appears in the air at a higher level than the observer. In either case, the images are inverted; and because of the irregular contour of the air layer, they are usually distorted. A somewhat different effect, known as *looming*, is produced by the refraction of light passing from rarefied air to a lower and denser layer. This results either in distortion, making distant objects appear grotesquely elongated vertically, or in lifting into view objects beyond the horizon. Looming effects are most frequently observed at sea.

Twilight and Afterglows

Twilight is that period between sunset and night and between night and sunrise. Civil Twilight is the period when the sun is between 0° and 6° below the horizon. Nautical Twilight is that period when the sun is between 0° and 12° below the horizon. Astronomical Twilight is that period when the sun is 0° and 18° below the horizon. Night is the period when the sun is more than 18° below the horizon.

The curtain of night is a relatively sharp slightly curved line across the sky that rises in the east in the evening. The night recedes in the east before the line of dawn, which is also a slightly curved line, appears. Both the curtain of night and the line of dawn are best observed from an aircraft flying at a high altitude from which position both lines can be sharply delineated.

Afterglow. A broad, high arch of radiance seen occasionally in the western sky above the highest clouds in deepening twilight. It is caused by the scattering effect exerted upon the components of white light by very fine particles of material suspended in the upper atmosphere. When used in this rather broad sense, the term embraces all the complex luminosities observed in the western twilight sky, but chiefly the purple light and the bright segment.

The purple light is the faint purple glow observed on clear days over a large region of the western sky after sunset, and over the eastern sky before sunrise. The purple light first appears, in the sunset case, for example, at a solar depression of 2°; at that time, it extends from about 35° to about 50° elevation above the solar point, and has an azimuthal extent of between 40° and 80°. Maximum intensity of the glow typically occurs at the time the sun is about 4° below the horizon. Increasing depression of the sun causes the top of the purple light to descend steadily toward the western horizon. The effect disappears at solar depression angles near 7°, being replaced in the western sky by the bright segment.

Whiteout. Also termed milky weather, whiteout is an atmospheric optical phenomenon of the polar regions in which the observer appears to be engulfed in a uniformly white glow. Neither shadows, horizon, nor clouds are discernible; senses of depth and orientation are lost; only very dark, nearby objects can be seen.

Aurora and air glow are described under **Aurora and Airglow.**

Polarization of Sky Radiation

There are three commonly detectable points of zero polarization of diffuse sky radiation, neutral points, lying along the vertical circle through the sun:

Arago Point. Named for its discoverer, the Arago point is customarily located at about 20° above the antisolar point; but it lies at higher altitudes in turbid air. The latter property makes the Arago distance a useful measure of atmospheric turbidity.

Babinet Point. This point typically lies only 15° to 20° above the sun, and hence is difficult to observe because of solar glare. The existence of this neutral point was discovered by Babinet in 1840.

Brewster Point. Discovered by Brewster in 1840, this neutral point is located about 15° to 20° directly below the sun; hence, it is difficult to observe because of the glare of the sun.

Zodiacal light is not an atmospheric phenomenon. See **Zodiacal Light.**

For references see entries on **Climate;** and **Meteorology.**

Additional Reading

Burke, W. L.: "Multiple Gravitational Imaging by Distributed Masses," *The Astrophysical J.*, Part 2, L1 (February 15, 1981).

do Carmo, M. P.: "Differential Geometry of Curves and Surfaces," Prentice-Hall, Englewood Cliffs, New Jersey, 1976.

Fodor, J.: "The Modularity of Mind," MIT Press, Cambridge, Massachusetts, 1983.

Fraser, A. B., and W. H. Mach: "Mirages," *Sci. Amer.*, 102–111 (January 1976).

Hirsch, M. W.: "Differential Topology," Springer-Verlag, New York, 1976.

Hoffman, D. D.: "The Interpretation of Visual Illusions," *Sci. Amer.*, 154–162 (December 1983).

Marr, D.: "Vision: A Computational Investigation into the Human Representation and Processing of Visual Information," Freeman, New York, 1982.

Meinel, A., and M. Meinel: "Sunsets, Twilights, and Evening Skies," Cambridge University Press, New York, 1983.

Tape, W.: "The Topology of Mirages," *Sci. Amer.*, 120–129 (June 1985).

Peter E. Kraght, Certified Consulting Meteorologist,
Mabank, Texas.

ATMOSPHERIC PRESSURE. Also termed *barometric pressure*, atmospheric pressure is the pressure exerted by the atmosphere as a consequence of gravitational attraction on the vertical column of air lying directly above the surface of the earth upon which the pressure is effective. As with any gas, atmospheric pressure is ultimately explainable in terms of the kinetic energy of impacting constituent atmospheric gases upon the surface that experiences the pressure.

Atmospheric pressure is one of several basic meteorological parameters. It is measured fundamentally by the height of a column of mercury (or other heavy fluids) in a sealed and evacuated tube, one end of which is exposed to the air. Atmospheric pressure forces the mercury to rise in the sealed and evacuated portion of the tube to a height at which the weight of the mercury exactly balances the weight of the air column resting on the open end of the tube. Such an instrument is called a barometer.

Air pressure is expressed in several ways. The most commonly used unit in meteorology is the *millibar* in which one millibar equals 1000 dynes per square centimeter. Atmospheric pressure averages about 1013.2 millibars at sea level. The kilopascal (kPa) is also a measure of atmospheric pressure. One kPa = 10 millibars. In the kPa system, average atmospheric pressure is 101.325 kPa. The height of the mercury column in a barometer is also used, either as millimeters or inches of mercury. Average sea level pressure in this system is 760 millimeters or 29.92 inches of mercury. Pounds per square inch is used in engineering. The term atmosphere is also used, one atmosphere being average sea level air pressure.

Hydrostatic Equation. Pressure-altitude relations in the atmosphere are mathematically precise and can be determined from the hydrostatic equation:

$$\frac{\partial p}{\partial h} = -\rho g$$

and the equation of state for air,

$$p = \rho R T$$

which leads to the relation

$$p = p_0 \left[\frac{T_0 - \lambda h}{T_0} \right]^{g/R\lambda}$$

where p = pressure
p_0 = pressure at height zero
h = altitude
ρ = density
g = acceleration of gravity
T = temperature (absolute)
T_0 = temperature at height zero (also absolute)
R = universal gas constant
λ = lapse rate and is constant

In the real atmosphere, many layers of air, each having its own approximately constant lapse rate of temperature, press down on each other to create the total air pressure at the base of the bottom layer. Air pressure results from

$$p_0 = p_1 + p_2 + p_3 \cdots + p_n = \sum_{1}^{n} p_i$$

where the p's are valid at the base of their respective layers. A fictitious lapse rate can be used that will yield nearly the same results as the combined pressures of the several uniform layers.

When the atmosphere is isothermal, as it very nearly is in the stratosphere, the pressure-height relationship becomes

$$p = p_0 e^{-gh/RT}$$

These relations state that (1) pressure decreases more rapidly with altitude when the temperature is low than when temperature is high; and (2) pressure decreases more rapidly with altitude when the lapse rate is large than when it is small.

The standard atmosphere is used for calibration of altimeters. It is not very often that the real atmosphere assumes the arbitrarily assigned values of the standard atmosphere; therefore, altimeters do not often indicate exact altitude.

Total mass of air per cubic centimeter is the *density of air*. It is given by the relation,

$$\text{density} = \frac{0.0012930}{1 + 0.00367t}\left[\frac{B - 0.378e}{760}\right]$$

where t = temperature, °C
B = barometric pressure expressed in mm of mercury
e = the partial pressure of water vapor in the air

Air density at standard conditions of 0°C and 760 millimeters of mercury is 0.0012930 gram per cubic centimeter of air free from water vapor. The standard density of air at 32°F and 14.7 pounds pressure is 0.081 pound per cubic foot (1.3 kilograms per cubic meter) and its composite molecular weight is 28.84.

The variation of air pressure with altitude is given in the accompanying table. See also **Atmosphere (Earth)**.

VARIATION OF AIR PRESSURE WITH ALTITUDE IN STANDARD ATMOSPHERE AS USED IN ALTIMETRY

ALTITUDE		PRESSURE		
Feet	Meters	Inches of Mercury	Millibars	Pounds per Square Inch
Sea Level	Sea Level	29.92	1013.2	14.7
1000	304.8	28.86	977.3	14.2
5000	1524	24.89	842.9	12.2
10,000	3048	20.58	696.9	10.1
15,000	4572	16.88	571.6	8.3
20,000	6096	13.75	465.6	6.8
25,000	7620	11.10	375.9	5.4
30,000	9144	8.88	300.7	4.4
40,000	12,192	5.54	187.6	2.7
50,000	15,240	3.44	116.5	1.7

Scale Height. This is a measure of the relationship between density and temperature at any point in an atmosphere; the thickness of a homogeneous atmosphere which would give the observed temperature:

$$h = kT/mg = RT/Mg$$

where k is the Boltzmann constant; T is the absolute temperature; m and M are the mean molecular mass and the molar mass of the layer; g is the acceleration of gravity; and R is the universal gas constant.

See also **Barometer**. For references see entries on **Climate;** and **Meteorology.**

Peter E. Knight, Certified Consulting Meteorologist,
Mabank, Texas.

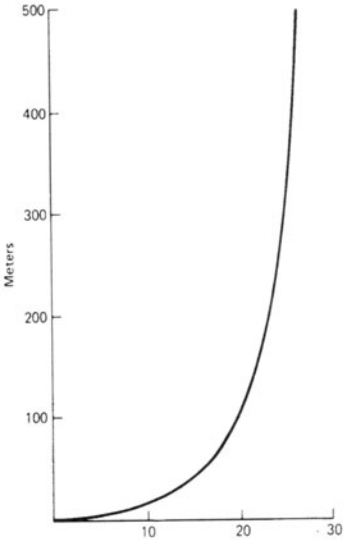
ATMOSPHERIC TURBULENCE. Air usually flows from one point to another in a turbulent manner, that is, the flow is infested with a multitude of small deviations of speed in all directions. This phenomenon has been observed for hundreds of years in the spreading and dissipation of smoke plumes. This can be demonstrated by inserting a pencil-sized smoke source into a wind stream. The thread of visible smoke does not extend downwind in a straight, thin line, but rather follows a zigzag path, spreading out and expanding downwind. This behavior demonstrates the presence of turbulence in the atmosphere.

When specific instantaneous speeds of the wind are measured at a point for a relatively large number of observations over a comparatively brief period, and this ensemble of data is averaged, the small deviations in speed all cancel each other with the mean speed remaining.

Moving air behaves as if embedded eddies of varying sizes roll and migrate, eventually to be absorbed into another part of the main airstream, merge with other eddies, or dissipate some distance from their origin. The result of turbulent flow is the transport of atmospheric pollutants, particulates, water vapor, heat, and momentum.

The *mixing length* is a rather unprecisely defined distance over which eddies of a certain size are able to transport their own embedded properties. Thus, during its life span, an eddy rolls and migrates from one point in the airstream to another, carrying its implanted pollutants, particulates, water vapor, heat, and momentum but shedding these as it travels to become totally lost at the end of its mixing length. The concept is similar to that of the mean free path in molecular theory.

Turbulent flow in the atmosphere has many significant meteorological consequences. Among others, the shape of the wind-speed profile just above the earth in the lower 500 meters depends in large part on the turbulent mixing in that layer. The speed profile is nearly always one in which there is a rapid increase just above the ground, but increasing less rapidly at higher levels. Exponential and logarithmic models describe the profile theoretically. See accompanying figure.



Wind speed profile based on the model: $V = 10 \log h$, where V is in knots and h is in meters.

Industrial pollutants are distributed upward and laterally by atmospheric turbulence as they are carried downwind in the large-scale airflow. Acid rain and snow fall hundreds of miles from the pollutant source region. A specialized field of industrial meteorology has emerged in the past 35 years to deal with the problems of industrial pollutants in the atmosphere.

Water vapor in the atmosphere originates primarily from the oceans, rivers, and lakes, as well as from transpiration from vegetation. These sources are all at the earth's surface. Eddies in the turbulent flow of the atmosphere distribute the water vapor and thus rain and snow fall far from the source of moisture.

Condensation and sublimation nuclei are dispersed extensively throughout the tropopause by turbulent flow. Salt particles from sea spray are among the nuclei that are distributed everywhere by atmospheric turbulence. These nuclei join other factors in causing rain and snow.

Heat is transported upward by eddies in turbulent flow, thus "cooling" the earth's surface and "warning" the higher atmospheric layers.

Eddies in turbulent flow of sufficiently large dimensions can alter the longitudinal air flow past aircraft in flight, thus causing rough-flying air. Most such rough air is simply an annoyance to travelers, but in some instances, may jolt the aircraft and toss items in the cabin around.

See also **Richardson Number;** and **Wind Shear.**

Peter E. Kraght, Certified Consulting Meteorologist,
Mabank, Texas.

ATOLL. A coral reef of ring-shape and appearing as a low, essentially circular, but sometimes elliptical or horseshoe shaped island. An atoll also may be a ring of closely spaced coral islets which encircle or nearly encircle a shallow lagoon where there is evidence of preexisting land of noncoral origin, and surrounded by deep water in the open sea. Atolls are common in the western and central Pacific Ocean.

ATOM. An atom is a basic structural unit of matter, being the smallest particle of an element that can enter into chemical combination. Each atomic species has characteristic physical and chemical properties that are determined by the number of constituent particles (protons, neutrons, and electrons) of which it is composed; especially important are the number Z of protons in the nucleus of each atom. To be electrically neutral the number of electrons in an atom must also be Z. The arrangement of these electrons in the internal structure of an atom determines its chemical properties. All atoms having the same atomic number Z have the same chemical properties, but differ in greater or lesser degree from atoms having any other value of Z. Thus, for example, all atoms of sodium ($Z = 11$) exhibit the same characteristic properties and undergo those reactions which chemists have found for the element sodium. Although these reactions are similar in some degree to those reactions characteristics of certain other elements, such as potassium and lithium, they are not exactly the same and hence can be distinguished chemically (see **Chemical Elements**), so sodium has properties distinctly different from those of all other elements. Individual atoms can usually combine with other atoms of either the same or another species to form molecules.

As explained in the entry on atomic structure, atoms having the same atomic number may differ in their neutron numbers or in their nuclear excitation energies.

The term *atom* has a long history, which goes back as far as the Greek philosopher Democritus. The concept of the atomic nature of matter was revived near the beginning of the nineteenth century. It was used to explain and correlate advancing knowledge of chemistry and to establish many of the basic principles of chemistry, even though conclusive experimental verification for the existence of atoms was not forthcoming until late in the nineteenth century. It was on the basis of this concept that Mendeleev first prepared a periodic table. See **Periodic Table of the Elements.**

Several qualifying terms are used commonly to refer to specific types of atoms. Examples of some of the terms are given in the following paragraphs.

An *excited atom* is an atom which possesses more energy than a normal atom of that species. The additional energy commonly affects the electrons surrounding the atomic nucleus, raising them to higher energy levels.

An *ionized atom* is an ion, which is an atom that has acquired an electric charge by gain or loss of electrons surrounding its nucleus.

A *labeled atom* is a tracer which can be detected easily, and which is introduced into a system to study a process or structure. The use of those labeled atoms is discussed at length in the entry **Isotope.**

A *neutral atom* is an atom which has no overall, or resultant, electric charge.

A *normal atom* is an atom which has no overall electric charge, and in which all the electrons surrounding the nucleus are at their lowest energy levels.

A *radiating atom* is an atom which is emitting radiation during the transition of one or more of its electrons from higher to lower energy states.

A *recoil atom* is an atom which undergoes a sudden change or reversal of its direction of motion as the result of the emission by it of a particle or radiation in a nuclear reaction.

A *stripped atom* is an atomic nucleus without surrounding electrons; also called a nuclear atom. It has, of course, a positive electric charge equal to the charge on its nucleus.

Subatomic particles and organization of the atom are discussed in the entry **Particles (Subatomic).**

ATOMIC CLOCKS. These devices make use of a property that is generally found only in systems of atomic dimensions. Such systems cannot contain arbitrary amounts of energy, but are restricted to an array of allowed energy values $E_0, E_1, \ldots, E_n$. If an atomic system is to change its energy between two allowed values, it must emit (or absorb) the energy difference—as by emission (or absorption) of a quantum of electromagnetic radiation. The frequency f_{ij} of this radiation is determined by the relation

$$|E_i - E_j| = \Delta E = hf_{ij}$$

where h is Planck's constant. The rate of an atomic clock is controlled by the frequency f_{ij} association with the transition from the state of energy E_i to the state of energy E_j of a specified atomic system, such as a cesium atom or an ammonia molecule. A high-frequency electromagnetic signal is stabilized in the atomic frequency f_{ij} and a frequency converter relates the frequency f_{ij} to a set of lower frequencies which then may be used to run a conventional electric clock.

The atomic frequency f_{ij} is, according to present knowledge, free of inherent errors. It is, in particular, not subject to "aging" since any transition which the system makes puts it in a state of completely different energy, where it cannot falsify the measurement. Herein lies the principal advantage over other methods of time measurement. Two atomic clocks have exactly the same calibration so long as they are calibrated against the same atomic transition. Atomic readings made in Boulder, Colorado and Neufchâtel, Switzerland between 1960 and 1963 differed on the average by less than 3 msec, whereas the deviation of the astronomically measured time TU_2 from atomic time is of the order of 50 msec. For this reason, the atomic second was adopted as the new time unit, by the Twelfth General Conference on Weights and Measures, in October 1964. This was initially defined as the time interval spanned by 9,192,631,770 cycles of the transition frequency between two hyperfine levels of the atom of cesium 133 undisturbed by external fields. See also "Atomic Clock" in entry on **Clock.**

The accuracy of present atomic clocks is limited by the thermal noises inherent at room temperatures. Theoretically, this limitation could be removed if the clocks were maintained in an atmosphere approaching absolute zero. However, some atomic clocks, like hydrogen maser clocks, stop oscillating when they are supercooled. In 1979, scientists at the Harvard-Smithsonian Center for Astrophysics overcame this problem by coating a supercooled maser cavity with carbon tetrafluoride. With the CF_4 frozen on the interior surfaces of the cavity, the oscillating hydrogen atoms could be reflected off the walls without becoming perturbed, thus preserving the phase of the oscillations. The researchers were able to keep a hydrogen maser clock operating at temperatures somewhat above 25 K. It has been estimated that a hydrogen maser clock cooled to about 25 K could run for 300 million years before losing one second of time, a factor some six times better than present hydrogen masers. The Center for Astrophysics is interested in improved maser clocks in connection with long-baseline interferometry and satellite tracking systems. It is also envisioned that a supercooled clock put on a space probe could be helpful in research on gravity waves and possibly provide clues toward better understanding the sun's mass distribution and angular momentum.

Refer to article on **Relativity and Relativity Theory** for description of the Sagnac effect. An atomic clock moved once around the Earth's equator in an easterly direction will *lag* a master clock at rest on the Earth by about 207.4 nanoseconds; a clock similarly moved in a westerly direction will *lead* the clock at rest by about 207.4 nanosceconds.

ATOMIC DISINTEGRATION. The name sometimes given to radioactive decay of an atomic nucleus and occasionally to the breakup of a compound nucleus formed during a nuclear reaction (see **Radioactivity**).

ATOMIC ENERGY. 1. The constitutive internal energy of the atom, which would be released when the atom is formed from its constituent particles, and absorbed when it is broken up into them. This is identical in magnitude with the total binding energy and is proportional to the mass defect. 2. Sometimes this term is used to denote the energy released as the result of the disintegration of atomic nuclei, particularly in large-scale processes, but such energy is more commonly called nuclear energy. See **Nuclear Power.**

ATOMIC ENERGY LEVELS. 1. The values of the energy corresponding to the stationary states of an isolated atom. 2. The set of stationary states in which an atom of a particular species may be found, including the ground state, or normal state, and the excited states.

ATOMIC FREQUENCY. The vibrational frequency of an atom, used particularly with respect to the solid state.

ATOMIC HEAT. The product of the gram-atomic weight of an element and its specific heat. The result is the atomic heat capacity per gram-atom. For many solid elements, the atomic heat capacity is very nearly the same, especially at higher temperatures and is approximately equal to $3R$, where R is the gas constant (Law of Dulong and Petit).

ATOMIC HEAT OF FORMATION. Of a substance, the difference between the enthalpy of one mole of that substance and the sum of the enthalpies of its constituent atoms at the same temperature; the reference state for the atoms is chosen as the gaseous state. The atomic heat of formation at 0 K is equal to the sum of all the bond energies of the molecule, or to the sum of all the dissociation energies involved in any scheme of step-by-step complete dissociation of the molecule.

ATOMIC HYDROGEN WELDING. Welding in a hydrogen atmosphere using heat from the arc between two tungsten electrodes. See also **Welding.**

ATOMIC MASS (Atomic Weight). As of the late 1980s, the current and internationally accepted unit for atomic mass is *1/12th of the mass of an atom of the* ^{12}C *nuclide and the official symbol is u.* The SI symbol u was selected so that it would indicate measurements made on the unified scale.[1]

It is interesting to note that prior to 1961, *two* atomic mass scales were used. Chemists preferred a scale based on the assignment of *exactly* 16, which experience had shown as the *average* mass of oxygen atoms as they are found in nature. On the other hand, physicists preferred to base the scale on a single isotope of oxygen, namely, ^{16}O (oxygen-16). The two scales differed because oxygen has three stable isotopes, ^{16}O, ^{17}O, and ^{18}O (as well as three identifiable radioactive isotopes, ^{14}O, ^{15}O, and ^{19}O).

Long before an understanding of the structure of the atom had been established and before the existence of isotopes was evidenced, several pioneers proposed what have become known as the concepts (laws) of:

> *Combining Volumes*—under comparable conditions of pressure and temperature, the volume ratios of gases involved in chemical reactions are simple whole numbers
>
> *Combining Weights*—if the weights of elements which combine with each other be called their 'combining weights,' then elements always combine either in the ratio of their combining weights or of simple multiples of these weights.

This then led to the establishment of the basic principle that the *combining weight of an element or radical is its atomic weight divided by its volume.*

Although the tables of atomic weights published today embrace all of the known chemical elements, it should be pointed out that the concept of combining weights stemmed exclusively from very early experiments strictly with *gases*. The kinetic theory of gases, which was developed from a line of logic that did not require the innermost understanding of the atom as we know it today, served as the early basis of how atoms react in quantitative proportions with each other to form compounds.

Boyle (1662) observed that at constant temperature the volume of a sample of gas varies inversely with pressure, but Boyle did not explain why this was so. Somewhat later, Charles (1787) refined the observation to the effect that the volume of any sample of a gas varies directly with the *absolute temperature* provided that the pressure is held constant. A few years later, Gay-Lussac (1808), in reporting the results of his experiments with reacting gases, observed that volumes of gases that are used or produced in a chemical reaction can be expressed in ratios of small whole numbers—a concept to become known as Gay-Lussac's law of combining volumes. It should be noted that the foregoing concepts proposed by Boyle, Charles, and Gay-Lussac were based upon experimental observations, not on theory.

An explanation for the law of combining volumes was given by Avogadro (1811) in which he proposed that equal volumes of all gases at the same pressure and temperature contain the same number of molecules. This, obviously, was an extension of Bernoulli's earlier thinking.[2] Avogadro's observations were essentially ignored and it remained for Cannizzaro (1858–1864) to develop, in a practical way, a method for computing the combining weights for gaseous compounds. This work led to the universal acceptance of Avogadro's principle. Cannizzaro used gas densities to assign atomic and molecular weights, basing his atomic weight scale on hydrogen. The hydrogen atom was assigned a value of 1 (approximately its assigned value today). The molecular weight of hydrogen was 2.

Much further research and careful experimentation was required to convey the principle to solid compounds. See entries on **Chemical Composition;** and **Chemical Formula.**

In returning to the attractive simplicity of combining weights in terms of ratios of small whole numbers, why then is it necessary, considering the standard for comparison ($^{12}C = 12$), to extend the atomic weight values to four and more decimal places? The principal answer is the presence of isotopes. Isotopes were unknown in the days of Boyle and other early pioneers. In essence, the atomic weight of an element is a *weighted average* of the atomic masses of the natural isotopes. The weighted average is determined by multiplying the atomic mass of each isotope by its fractional abundance and adding the values thus obtained. A fractional abundance is the decimal equivalent of the percent abundance. However, for the standard of comparison, obviously a specific isotope was selected rather than a weighted average for that element.

One might also query—since we know so much today about the masses of the protons, neutrons, and electrons comprising an atom,[3] why not simply add up these specific values for a given atom? This, of course, still would not relieve the isotope problem, but it is not accurate to do so mainly because of Einstein's equation $E = mc^2$. As pointed out by Mortimer, with the exception of ^1_1H, the sum of the masses of the particles that make up a nucleus will always differ from the actual mass of the nucleus. If the required nucleons were brought together to form a nucleus, some of their mass would be converted into energy. Called the *binding energy*, this is also the amount of energy required to pull the nucleus apart.

For practical purposes, a majority of elements have a constant mixture of natural isotopes. For example, mass spectrometric studies of chlorine show that the element consists of 75.53% $^{35}_{17}Cl$ atoms (mass = 34.97 u) and 24.47% $^{37}_{17}Cl$ atoms (mass = 36.95 u). Experience has shown that any sample of chlorine from a natural source will consist of these two isotopes in this proportion.

ATOMIC MASS UNIT. See **Units and Standards.**

ATOMIC NUMBER. See **Chemical Elements.**

ATOMIC PERCENT. The percent by atom fraction of a given element in a mixture of two or more elements.

---

[1] It should be stressed that u, as the standard for comparing the masses (weights) of all chemical elements in all kinds of chemical compounds, refers not simply to the carbon atom, but rather to one very specific isotope, carbon-12. There are two stable isotopes of carbon, ^{12}C and ^{13}C, and four known radioactive isotopes, ^{10}C, ^{11}C, ^{14}C, and ^{15}C.

[2] An attempt was made by Daniel Bernoulli (1738) to explain Boyle's law on the basis of what later became known as the *kinetic theory of gases*. Bernoulli introduced the concept that the pressure of a gas results from the collisions of gas molecules within the walls of the gas container. This established a connection between the *numbers* of gas molecules present and their kinetic energy present at any given temperature.

[3] Mass of proton is 1.007277 u; of neutron is 1.008665 u; mass of electron is 0.0006486 u.

ATOMIC PLANE. A plane passed through the atoms of a crystal space lattice, in accordance with certain rules relating its position to the crystallographic axes. See **Mineralogy**.

ATOMIC RADIUS. See **Chemical Elements.**

ATOMIC SPECIES. A distinctive type of atom. The basis of differentiation between atoms is (1) mass, (2) atomic number, or number of positive nuclear charges, (3) nuclear excitation energy. The reason for recognizing this third class is because certain atoms are known, chiefly among those obtained by artificial transmutation, which have the same atomic (isotopic) mass and atomic number, but differ in energetics.

ATOMIC SPECTRA. An atomic spectrum is the spectrum of radiation emitted by an excited atom, due to changes within the atom; in contrast to radiation arising from changes in the condition of a molecule. Such spectra are characterized by more or less sharply defined "lines," corresponding to pronounced maxima at certain frequencies or wavelengths, and representing radiation quanta of definite energy.

The lines are not spaced at random. In the spectrum of hydrogen, for example, there is a prominent red line (H_α) and, far from it, another (H_β) in the greenish-blue, then after a shorter wavelength interval a blue-violet line (H_γ), and after a still shorter interval another violet line (H_δ), etc. One has only to plot the frequencies of these lines as a function of their ordinal number in the sequence, to get a smooth curve which shows that they are spaced in accordance with some law. In 1885, Balmer studied these lines, now called the Balmer series, and arrived at an empirical formula which in modern notation reads

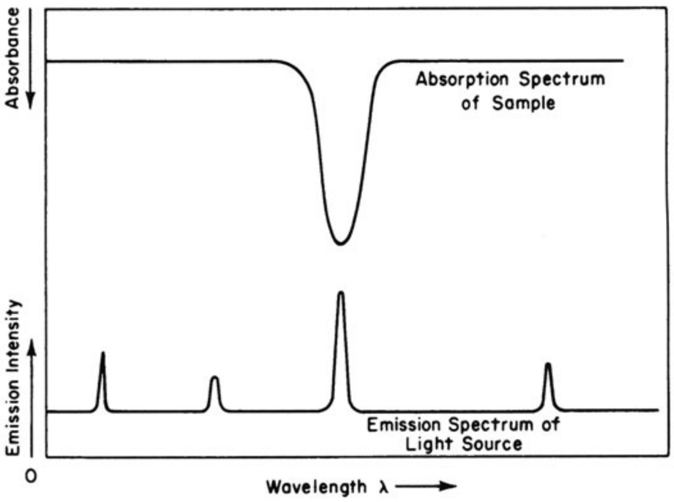
$$v = Rc\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right)$$

It gives the frequency of successive lines in the Balmer series if R is the Rydberg constant, c the velocity of light, $n_1 = 2$, $n_2 = 3, 4, 5, \ldots$. As n_2 becomes large, the lines become closer together and eventually reach the series limit of $v = Rc/4$. Ritz, as well as Rydberg, suggested that other series might occur where n_1 has other integral values. These, with their discoverers and the spectral region in which they occur are as follows:

Lyman series, far ultraviolet, $n_2 = 2, 3, 4, \ldots, n_1 = 1$
Paschen series, far infrared, $n_2 = 4, 5, 6, \ldots, n_1 = 3$
Brackett series, far infrared, $n_2 = 5, 6, 7, \ldots, n_1 = 4$
Pfund series, far infrared, $n_2 = 6, 7, 8, \ldots, n_1 = 5$

See also **Energy Level.**

ATOMIC SPECTROSCOPY. Chemical analysis by atomic absorption spectrometry involves converting the sample, at least partially, into an atomic vapor and measuring the absorbance of this atomic vapor at a selected wavelength which is characteristic for each element. The measured absorbance is proportional to concentration, and analyses are made by comparing this absorbance with that given under the same experimental conditions by reference samples of known composition. Several methods of vaporizing solids directly can be used in analytical applications. One of the first methods used was spraying a solution of the sample into a flame, giving rise to the term "absorption flame photometry." When a flame is used, the atomic absorption lines are usually so narrow (less than 0.05 Å) that a simple monochromator is not sufficient to obtain the desired resolution. Commercial atomic absorption spectrophotometers overcome this difficulty by using light sources which emit atomic spectral lines of the element to be determined under conditions which ensure that the lines in the spectrum are narrow, compared with the absorption line to be measured. With this arrangement, peak absorption can be measured, and the monochromator functions only to isolate the line to be measured from all other lines in the spectrum of the light source. See accompanying figure.



Lines emitted by the light source are much narrower than the absorption line to be measured.

Atomic spectra, which historically contributed extensively to the development of the theory of the structure of the atom and led to the discovery of the electron and nuclear spin, provide a method of measuring ionization potentials, a method for rapid and sensitive qualitative and quantitative analysis, and data for the determination of the dissociation energy of a diatomic molecule. Information about the type of coupling of electron spin and orbital momenta in the atom can be obtained with an applied magnetic field. Atomic spectra may be used to obtain information about certain regions of interstellar space from the microwave frequency emission by hydrogen and to examine discharges in thermonuclear reactions.

ATOMIZATION. The breaking-up of a liquid into small droplets, usually in a high-speed jet or film.

ATRIUM. Literally, an entrance chamber, and so applied to various organs. 1. The main part of the cavity of the middle ear. 2. The vestibule of the female genital passages. 3. A chamber into which the genital organs open in the flatworms. 4. A cavity formed of folds of the body wall in *Amphioxus* and the tunicates, which partially surrounds the pharynx and opens to the exterior by an atriopore. 5. The chamber at the end of an air tube in the lungs, with which the ultimate air sacs or alveoli communicate. 6. The chamber of the heart in vertebrates which empties into the ventricle. In this sense the term atrium is frequently replaced by auricle, although in strict terminology the auricle refers only to a small appendage of the atrium.

ATROPHY. Physiologic or pathologic reduction in size of a mature cell or organ, usually with some degree of degeneration. Following certain diseases, types of accidents and surgery where nerves may be damaged or cut, atrophy may be temporary or permanent. Where there is traumatic injury of the nerves that cannot be repaired, atrophy is progressive. Atrophy or degeneration of the anterior lobe of the pituitary gland in adults results in Simmonds-Sheehan disease, a disorder characterized by an extreme appearance of aging. The metabolic functions of the body are affected and eventually mental functions decline as well. Pituitary atrophy is believed to result from anoxia, a lack of oxygen reaching the gland after a condition, such as postpartum hemorrhage, where excessive blood is lost. Simmonds-Sheehan disease sometimes is confused with *anorexia nervosa*, a serious nervous condition in which the patient eats little food and is greatly emaciated. A metabolic test (Metapirone test) can distinguish the two conditions. Further, in females, amenorrhea (absence of menstruation) is a constant feature of Simmonds-Sheehan disease.

ATTENUATION. 1. In its most general sense, attenuation is reduction in concentration, density or effectiveness. 2. In psychological statistics, the weakening of the correlation between two variables due to errors of measurement on them. 3. In radiation theory, attenuation is used to express the reduction in flux density, or power per unit area, with distance from the source; the reduction being due to absorption and/or scattering. In this usage, attenuation does not include the inverse-square decrease of intensity of radiation with distance from the source. 4. The same restriction applies to the use of the term in nuclear physics, where attenuation is the reduction in the intensity of radiation on passage through matter where the effect is usually due to absorption and scattering. 5. In an electric network or line, attenuation is loss, usually of current. See **Attenuation Factor; Attenuator.**

In terms of scientific instruments, the Scientific Apparatus Makers Association defines attenuation as: (1) A decrease in signal magnitude between two points, or between two frequencies; and (2) The reciprocal of gain, when the gain is less than one. Attenuation may be expressed as a dimensionless ratio, scalar ratio, or in decibels as 20 times the $\log_{10}$ of that ratio.

ATTENUATION FACTOR. 1. A measure of the opacity of a layer of material for radiation traversing it. It is equal to I_0/I, in which I_0 and I are the intensities of the incident and emergent radiation, respectively. In the usual sense of exponential absorption

$$I = I_0 e^{-\mu x}$$

where x is the thickness of the material and μ is the absorption coefficient. 2. A meaning similar to that in (1) is current in electrical circuit applications, where the attenuation factor is the ratio of the input current to the output current of a line or network.

ATTENUATION RATIO. The magnitude of the propagation constant.

ATTENUATION (Sideband). That form of attenuation in which the transmitted relative amplitude of some component(s) of a modulated signal (excluding the carrier) is smaller than that produced by the modulation process.

ATTENUATOR. The attenuator, often called a pad, is a network designed to introduce a definite loss in a circuit. It is designed so the impedance of the attenuator will match the impedance of the circuit to which it is connected, often being connected between two circuits of different impedance and serving as a matching network as well as an attenuator. It is distinguished from a simple resistance in that the impedance of an attenuator does not change for various values of its attenuation. It is a valuable unit in making many laboratory tests on communications equipment, where it is used to adjust the outputs of two pieces of apparatus or for two different conditions so the relative merits may be determined from the attenuator setting. In much communication work it is desirable to transmit power at a higher level than will be used in order to overcome circuit noises, and then to reduce it to the proper value at the receiving end by a pad. It is usually calibrated in decibels and thus indicates the attenuation introduced by it.

Among the types of attenuators, there is the *coaxial line attenuator*, which, as its name indicates, is designed for use in a coaxial line. It may be fixed or variable. One of its special types is the *chimney attenuator*, which received its name from the appearance of the stub lines. The *flap or fin attenuator* is a waveguide attenuator in which a flap or fin of conducting materials is moved into the guide in such a manner as to cause power absorption. The *transverse film attenuator* consists of a conducting film placed transverse to the axis of a waveguide.

ATTITUDE. The position or orientation of an aircraft or spacecraft, either in motion or at rest, as determined by the relationship between its axes and some reference line or plane or some fixed system of reference axes.

ATTRITION (Geology). From the Latin *attritio* meaning a grinding or rubbing down, and used in the terminology of geological science to refer to the grinding of particles through the transporting power of wind, running water, or by the movement of glaciers.

ATTRITION MILLS. Equipment of this type is used in the process industries to reduce the size of various feeds. Attrition connotes a rubbing action, although this action usually is combined with other forces, including shear and impact. Attrition mills also are referred to as disk mills and normally comprise two vertical disks mounted on horizontal shafts, with adjustable clearance between the vertical disks. In some designs, one vertical disk may be stationary. In other designs, the two vertical disks rotate at differential speeds or in opposite directions. Material is fed to the mill so that it is subjected to a tearing or shredding action. Because of the frictional nature of the operation, temperatures build up and heat-sensitive materials cannot be size-reduced in this type of equipment. Attrition mills sometimes are used principally as mixers to provide an intimate blending of powders. Special plates are used to permit intensive blending with a minimum of grinding action. Throughput rates per horsepower required are high.

AUDIBILITY. The wide loudness range of the human ear is exemplified by the fact that the most intense sound that can be tolerated is a million million times greater in intensity than a sound that is just audible. This is a range of approximately 120 decibels. The decibel scale is a logarithmic ratio scale. The frequency range (*audio frequency*) of hearing is usually stated as 16 Hz to 20,000 Hz. The ear is most sensitive in the middle-frequency range of 1,000 to 6,000 Hz. Few individuals can hear above 20,000 Hz. Below 15 Hz, if detected, the sound normally is perceived not as a note, but as individual pulses.

In terms of discrimination of frequency and intensity, it is possible for about 1,400 pitches and 280 intensity levels to be distinguished. The rather phenomenal aspects of hearing can be observed in such behavior as localization of sounds (*auditory localization*), speech perception and, in particular, the understanding of one voice in the noisy environments of many. Acoustic events that last only a few milliseconds also can usually be detected.

The instrument for measuring hearing acuity is termed an *audiometer*. See also **Auditory Organs; Hearing and the Ear.**

AUDIO FREQUENCY PEAK LIMITER. A circuit used in an audio frequency system to cut off signal peaks that exceed a predetermined value.

AUDIOGRAM. A graph showing the hearing loss, the percent of hearing loss, or the percent of hearing as a function of frequency. See also **Hearing and the Ear.**

AUDITORY ORGANS. Organs sensitive to stimulation by sound waves. True auditory organs occur in arthropods and vertebrates. In the former they vary considerably but in the latter they are the ears and can be traced through their variations to a common structural foundation.

The simplest arthropod auditory organ is known as a chordotonal organ. It consists of a nerve ending with accessory cells connected with the body wall, which is apparently the immediate source of the vibrations to which the organ responds. More elaborate auditory organs are found in grasshoppers, katydids, mosquitoes, and related species. In the grasshoppers they are located on the sides of the first abdominal segment, in the katydids in the front tibiae, and in the mosquitoes at the base of the antennae. In all forms the scolophore is the essential sensory ending; accessory structures vary to a greater degree but usually include a modification of the cuticula which serves as a resonating membrane, or tympanum.

The essential auditory portion of the vertebrate ear is the cochlea, a spiral organ of elaborate structure containing terminations of the auditory nerve. This organ is part of the inner ear. In the mammals the outer ear includes the pinna, usually called the ear, and the external auditory canal leading inward to the tympanum or ear drum which vibrates in response to sound waves. Between these two regions lies the cavity of the middle ear, derived from the pharynx and connected with it by the Eustachian tube. The middle ear is bridged by a series of small bones, the hammer, anvil, and stirrup, which convey the vibrations of the tympanum mechanically to the liquid in the inner ear. These parts are variably developed in vertebrates below the mammals, all of which have simpler ears than described. The ears of bats play a unique part in the avoidance of obstacles during flight.

See also **Hearing and the Ear; Sensory Organs.**

AUGEN-GNEISS. A gneissoid rock that contains lenticular crystals or mineral aggregates resembling "eyes." Derived from the German *augen*, eyes.

AUGER EFFECT. A process, discovered by P. Auger, in which the energy released in the de-excitation of an excited electronic energy state of an atom is given to another one of the bound electrons rather than being released as a photon. This type of transition is usually described as radiationless. The process usually occurs only for transitions in the x-ray region of energy states. The final state corresponds to one higher degree of ionization than does the initial state. The ejected electron has kinetic energy equal to the difference between the energy of the x-ray photon of the corresponding radiative transition and the binding energy of the ejected electron.

AUGHEY SPARK CHAMBER. A device in which an electrical spark is produced between two hollow tubular electrodes. A gas passed through the electrodes into the space where the spark occurs will be excited and caused to emit its characteristic spectrum. Traces of contamination in the gas, such as dust particles, will be efficiently excited so that their spectrum may be observed.

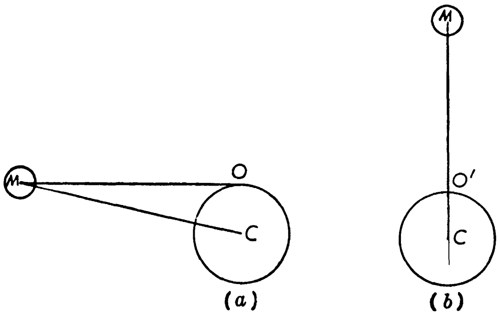
AUGITE. This mineral is a common monoclinic variety of pyroxene whose name is derived from the Greek word meaning "luster," in reference to its shining cleavage faces. Chemically it is a complex metasilicate of calcium, magnesium, iron and aluminum. Color, dark green to black, may be brown or even white; hardness, 5–6; specific gravity, 3.23–3.52. Augite is important as a primary mineral in the igneous rocks and also as secondary mineral. The white augite is called leucaugite from the Greek word meaning "white." Chemical analysis reveals this variety as containing little or no iron. Augite is of widespread occurrence.

See also **Pyroxene.**

AUGMENTATION. A term used by astronomers to indicate the increase in apparent diameter of the moon, or of any other object close enough to the earth to be observed as a disk, as the altitude of the object increases.

In Fig. (a), we have a representation of conditions for the object, *M*, on the horizon for an observer, *O*; in Fig. (b), the object, *M*, is at the zenith on the meridian for the observer, *O′*. In both figures, *C* represents the center of the earth. The distance, *CM*, of the object from the center of the earth is assumed to be a constant. Examination of the figures will show at once that (a) gives the maximum distance of the object from the observer, whereas (b) gives the minimum value of this distance. Since apparent angular diameter of an object increases with decrease of distance, and since apparent size is defined as the apparent angular diameter, the object is seemingly larger under conditions (b) than (a), i.e., larger on the meridian than when rising.

For the sun and planets, augmentation is too small to be considered except in the most refined observations of the altitude of a limb. However, in the case of the moon, augmentation may amount to as much as 37″. Failure to properly correct for this effect, when a limb of the moon is observed for determination of a line of position in navigation, might introduce an error as great as 0.3 mile in the position of a ship.



(a) (b)

Demonstration of augmentation.

AUREOLE (Geology). The contact metamorphic zone of varying width that often surrounds an igneous intrusion. Such areas of contact metamorphism often contain valuable ore deposits, especially when they surround batholiths which have intruded sedimentary formations.

AURIC AND AUROUS. Prefixes often used in the naming of gold salts of valence +3 (ic) and +1 (ous). Thus, auric chloride, aurous nitrate, and so on.

AURIGA (the charioteer). This constellation is best known because it contains the bright star Capella (the she-goat) (α Aurigae) and her kids. The kids are three fainter stars, forming, to the naked eye, a small triangle, which always serves to distinguish Capella from other bright stars on a clear night. Capella is a bright star, yellowish in appearance, and of the same spectral type as our sun. The star, however, is so much larger than our sun that, in spite of its great distance (49 light-years), it appears as first magnitude, whereas the sun at the same distance would be sixth magnitude, or barely visible to the naked eye on a clear moonless night. Capella A is a spectroscopic binary with a period of 104 days. More distant are two additional M dwarf components, making Capella a four-star system, or small cluster. See map accompanying the entry **Constellations.**

AURORA AND AIRGLOW. The visual aurora consists of luminous forms (arcs, rays, bands) in the night sky, usually confined to high latitudes and based in the ionospheric E region. See also **Ionosphere.** The airglow consists of a faint relatively uniform luminosity which is worldwide in occurrence and, except under exceptional conditions, can only be observed instrumentally. The distinction between faint aurora and bright airglow in auroral regions is not clear.

The luminosity arises from emissions of the atmospheric constituents in the atomic, molecular or ionized forms. The chief emissions in the visible region, with approximate intensities in Rayleighs for a bright aurora and temperate latitude airglow, are shown in the accompanying table. There are many other emissions in the infrared and ultraviolet. In bright aurorae, the colors can be seen visually; faint aurorae appear grayish white since they approach the color vision threshold.

An auroral arc is a narrow horizontal band of light up to hundreds of kilometers long (usually geomagnetic east-west). The term arc derives from its appearance from the earth's surface due to perspective. A band is a portion of an arc showing, distortion normal to its length. Auroral rays have been likened to searchlight beams; they lie along the geomagnetic field direction and may be several hundred kilometers long. Arcs and bands may be homogeneous or rayed. Particularly dramatic displays of the aurora are shown in Figs. 1 through 3.

Isolines of auroral occurrence are approximately centered on the geomagnetic poles. The auroral zones are defined as the regions of maximum occurrence. They are roughly circular with a radius of approximately 23° of latitude. The northern auroral zone reaches its lowest geographic latitude over eastern Canada; the southern, over the ocean south of Australia. At times of geomagnetic disturbance, the aurora appears at lower latitudes and in very great magnetic storms may be observed in the tropics. The frequency of occurrence of aurorae at lower latitudes correlates with the cycle of solar activity.

Within recent years, it has been found that a relatively uniform auroral glow exists over the polar cap, extending through and beyond the classical auroral zone, on which auroral forms appear as bright patches which are visible merely because of contrast with their surroundings. Another relatively recent finding is the thesis of a local time dependence in the daily maximum of auroral occurrence which is at about 68° geomagnetic at midnight and 75 to 80° at noon. An inner auroral zone at 75 to 80° geomagnetic could possibly explain the observed results.

Many auroral forms are probably caused by the precipitation of particles (mainly electrons) into the ionosphere. Their origin is obscure, but studies suggest that they are derived from the outer regions of the magnetosphere, and are accelerated and precipitated in an irregular

CHIEF AURORAL AND AIRGLOW EMISSIONS IN THE VISIBLE REGION[a]

Emission	Spectral Region or Wavelength	Approximate Height (Kilometers)	Approximate Intensity (Rayleighs)	
			Bright Aurora[b]	Nightglow
Ol	5577Å	90–110	100.000	250
	6300, 6364Å	160	>50.000	150
NII	Blue to red		25.000	
H (Balmer Series)	Red, blue	E-layer	1.000	
N_2 (1st Positive)	Red	D-layer	50.000	
(2nd Positive)	Violet	D-layer	100.000	
N_2^+ (1st Negative)	Blue-violet		165.000	
O_2^+ (1st Negative)	Red-yellow	D-layer	10.000	
Nal	5890, 5896Å	80–90		100 (winter) 20 (summer)
OH	Red-yellow	60–100		100
O_2 (Herzberg)	Blue-violet	90–100		15

[a] Adapted from Chamberlain: "Physics of the Aurora and Airglow." Academic Press, New York, 1961, to which reference should be made for data including ultraviolet and infrared emissions.

[b] International Brightness Coefficient III (brightness of moonlit cumulus clouds).

Note: Emissions are highly variable or absent with type and latitude of aurora. Heights are given only when well-defined.

manner on the high latitude side of the outer radiation belt through some mechanism (e.g., turbulence) which is probably related to the solar wind. It is doubtful if precipitation of trapped particles from the outer radiation belt causes aurora directly, except in great magnetic storms.

A strong ionospheric current system is seated approximately in the classical auroral zone, but the detailed relation between aurorae and the electric currents is obscure.

The so-called radio aurora signifies the ionization in the E-layer that is associated with magnetic disturbances, and gives rise to characteristic type radio reflections in the VHF (30 to 300 MHz) band and less often at higher frequencies. It has been suggested that radio aurora may be identified with the optical aurora, but little evidence exists for this.

The chief characteristic of the ionization is that it is aligned along the earth's magnetic field, the size of the irregularities ranging from meters to kilometers in length. The mechanism producing it is obscure; wind shears and particle precipitation probably contribute.

The pattern of ionization usually shows a systematic movement which in and below the auroral zone is statistically very similar to the ionospheric disturbance current system, but there are difficulties in interpreting the movement as that of the electrons in the current system. Other interpretations are that the movement is that of the ionizing sources, or even sound waves.

Frequency electromagnetic noise emission (hiss), centered around 8 kHz is observed in association with aurora. Satellite observations have sometimes shown correspondence between electron precipitation, auroral light intensity, and hiss, but at other times the correlation is poor. Theories of this noise emission all consider the interaction of a stream of particles with the surrounding plasma. Traveling wave tube amplification, Cerenkov radiation, or Doppler-shifted cyclotron generation by protons have been suggested.

The airglow is subdivided into the dayglow, twilightglow, and nightglow. The sodium intensity in the nightglow and twilightglow is highest

Fig. 1. A beautifully looped curtain aurora over Alaska. (*Geophysical Institute, University of Alaska.*)

Fig. 2. Auroral draperies. The drapery at the left is seen nearly edge-on. (*Hessler: Chamberlain, "Physics of the Aurora and Airglow," Academic Press.*)

Fig. 3. Homogeneous horseshoe band. (*Hessler: Chamberlain, "Physics of the Aurora and Airglow," Academic Press.*)

in local winter, but seasonal and diurnal variations of the other emissions are not clear as there are marked latitude effects and distinct patchiness. The origin of the nightglow is obscure, though an important part of the oxygen red emission is excited by electron-ion recombination in the F layer. At the 85 to 100 kilometer level, there are complex chemical reactions involving oxides of nitrogen as well as the free gases and ions. The energy sources are far from understood; winds, turbulence, the quiet day ionospheric current system, thermal excitation, and even particles may contribute.

Considerable knowledge pertaining to auroral activity is being gained from photographs taken from satellites. The most illuminating aspects of recent pictures are the large field of view covering a substantial fraction of the auroral oval and coverage of formerly inaccessible areas. Feldstein and Starkov (see references) have suggested that auroral activity occurs along an oval that surrounds the north geomagnetic pole and along a similar oval around the south geomagnetic pole. The position of the oval varies with geomagnetic activity. Its geomagnetic colatitudes are about 23° on the night side and 15° on the day side during periods of moderate geomagnetic activity. Observations indicate that the aurora frequently displays an eddy-like form with a characteristic length of a few hundred kilometers. Hasegawa (see references) has suggested that this may be the result of kink instability in the field-aligned sheet current proposed by Akasofu and Meng (see references). A short summary of auroral photographic studies conducted by U.S. Air Force Weather Service satellites is given in *Science*, **183**, 4128, 951–952 (1974).

In northern latitudes, auroral displays are called *aurora borealis, aurora polaris*, or *northern lights*. In southern latitudes, they are called *aurora australis*.

AUSCULTATION. This term is applied to the examination of the sounds within the chest, abdomen, heart, or larger blood vessels. It is carried out by listening with a stethoscope, or by applying the ear directly to the surface of the body. See also **Stethoscope.**

AUSTENITE. The solid solution based upon the face-centered cubic form of iron. The most important solute is usually carbon, but other elements may also be dissolved in the austenite. See also **Iron Metals, Alloys, and Steels.**

AUTHIGENOUS (or Authigenic). A geologic term proposed by Kalkovsky in 1880, meaning generated on the spot, and referring particularly to the primary and secondary minerals of igneous rocks and the cements of sedimentary rocks.

AUTOCHTHONOUS. A geologic term proposed in 1888 for sedimentary rocks which have been formed in place. Now generally used to designate bedrock masses that have remained in place in a mountain belt, the term allochthonous denoting masses that have been moved long distances.

AUTOCLASTIC. A term proposed by Van Hise in 1894 for crush breccias or fault breccias which have been fragmented in place.

AUTOCOLLIMATOR. 1. A device by which a lens makes diverging light from a slit parallel, and then after the parallel light has passed through a prism to a mirror and been reflected back through the prism, the same lens brings the light to a focus at an exit slit. 2. A telescope provided with a reticle so graduated that angles subtended by distant objects may be read directly. 3. A convex mirror placed at the focus of the principal mirror of a reflecting telescope and of such curvature that the light after reflection leaves the telescope as a parallel beam.

AUTOCORRELATION. The correlation of the members of a time-series among themselves.

Autocorrelation Coefficient. If ξ_t is a stationary stochastic process with mean m and variance σ^2 the autocorrelation coefficient of order k is defined by

$$\rho_k = \rho_{-k} = \frac{1}{\sigma^2} E(\xi_t - m)(\xi_{t+k} - m)$$

where the expectation relates to the joint distribution of ξ_t and ξ_{t+k}.

In a slightly more limited sense, if x_t is the realization of a stationary process with mean m and variance σ^2 the autocorrelation coefficients are given by a similar formula where the expectation is to be interpreted as

$$\lim_{n_2 - n_1 \to \infty} \frac{1}{(n_2 - n_1)} \sum_{j=n_1}^{n_2} (x_{t+j} - m)(x_{t+j+k} - m)$$

The expression is also applied to the correlations of a finite length of the realization of a series. Terminology on the subject is not standardized and some writers refer to the latter concept as serial correlation, preferring to denote the sample value by the Latin derivative "serial" and retaining the Greek derivative "auto" for the whole realization of infinite extent.

Autocorrelation Function The graph of the autocorrelation coefficient as ordinate against the order k as abscissa is called the correlogram. When the series is continuous in time, the set of auto coefficients may be summarized in an autocorrelation function. This is the autocovariance divided by the variance, e.g., for a series with zero mean and range $a \leq t \leq b$, defined at each time point, is given by

$$\rho(\tau) = \frac{1}{b - t - a} \int_a^{b-t} u(t)u(t+\tau)\, dt \Big/ \frac{1}{a-b} \int_a^b u^2(t)\, dt$$

The limits a and b may be infinite subject to the existence of the integrals or sums involved.

The numerator of this expression is called the autocovariance function.

Sir Maurice Kendall, International Statistical Institute, London.

AUTOGAMY. A process of nuclear reorganization in protozoa in which the nucleus divides, each half undergoes a maturation, and the two persisting functional nuclei reunite. In the modified process known as paedogamy the individual forms a cyst within which it divides into two cells which reunite after the nuclear transformation is completed.

AUTOGENOUS. Self-generated, originating within the body. The term is usually applied to vaccines that are made from a patient's own bacteria as opposed to stock vaccines which are made from cultures grown from standard strains.

AUTOGIRO. See **Helicopters and V/STOL Craft.**

AUTOIMMUNITY. See **Immune System and Immunology.**

AUTOINTOXICATION. Seldomly used term to describe poisoning by a substance generated within the body and which the body is unable to eliminate without treatment.

AUTOIONIZATION (or Preionization). Some bound states of atoms have energies greater than the ionization energy. An atom which is in a discrete energy state above the ionization point can ionize itself automatically with no change in its angular momentum vectors if there is a continuum with exactly the same characteristics. This process is called autoionization.

AUTOLYSIS. The energy derived from biological oxidations in living cells serves to promote anabolic processes, i.e., to produce relatively complex, highly ordered molecules and structures, and thus normally keeps living cells in a steady state remote from equilibrium. In organisms that lack cellular nutrients or oxygen (or in dead organisms or cells that have been disrupted so as to destroy much subcellular organization), the opposing catabolic tendency toward equilibrium, including the tendency toward degradation of macromolecules to simpler monomeric subunits, is not counterbalanced. These degradative processes, many of them enzymatically catalyzed, are collectively termed autolysis. Autolytic processes may include, for example, hydrolysis of proteins catalyzed by proteolytic enzymes or hydrolysis of nucleic acids catalyzed by nucleases. Autolysis of tissues (e.g., liver homogenate) has sometimes been used as a method for releasing bound molecules (e.g., vitamins or coenzymes) into free soluble form.

AUTOMATION. Possibly more aptly termed *automation engineering*, automation is a design engineering philosophy that is directed toward enchancing the automatic nature (sometimes called automaticity) of a machine, process, or other type of work system.[1] The objective of automation is to cause the work system to be as automatic, i.e., self-acting, self-regulating, and self-reliant, as may be possible—but against the real and practical backdrop of various economic, environmental, social, and other restraints. Because of these restraints, the work systems encountered on a day-to-day basis are only *partially automated*.

One definition of automation[2] was proposed in 1947 as "the automatic handling of workpieces into, between, and out of machines." As viewed in the late 1980s, this is a limited definition, although still accurate as far as it goes. Some authorities claim that automation is a contraction of the more-difficult-to-say word, *automization*. Still other scholars claim that automation was coined from *auto*matic and opera*tion*.

As pointed out by Mumford,[3] the curse of labor was described by the early Egyptians, who mentioned the daily hardships, the filth, the danger, the nightly weariness of producing goods. Later, the oppression of labor was recognized by the Greeks in the fifth century B.C. and by the Florentines in the twelfth century A.D. Prior to the last century or two, earlier people tended to look toward a force (leading to the modern concept of automation) that would abolish all work and, as described by Mumford, "the most desirable life possible would be one in which magical mechanisms or robots would perform all the necessary motions under their own power, without human participation of any kind. In short, the idea of the mechanical *automation*, which would obey all orders and do all the work." Thus, the negative connotations of automation in terms of adverse effects on the economy of a human work force did not arise seriously until the present century.

Numerous scientific and engineering disciplines make up the technical foundation for automation. Very prominent are electronics, electrical, mechanical, chemical, metallurgical, and industrial engineering; measurement and control technology, computer, information, and communication sciences—all supported by the principles of physics and mathematics.

Advantages/Limitations of Automation

As is apparent from the numerous technical articles in this encyclopedia, advanced automated systems are available today and further advances seem close at hand. Thus, a former question, "Is automation possible?" has been displaced by the query, "Is automation profitable?" As is essentially true of all business concerns, automation is welcomed most where it contributes to profit. Of the several dividends yielded by manufacturing processing automation, two are uppermost—improved productivity and better product quality.

1. *Improved productivity* of machines and people is a dividend that almost always translates into greater profitability and return on investment. Several factors enter into improved productivity, but two are most important:
   (a) *Increased production capacity*—more goods produced per manufacturing floor area, machines installed, and the human work force. In terms of machines, automation usually increases the duty cycle for machines, thus yielding more machine hours per day.
   (b) *Better inventory control* (flow of materials and energy throughout the plant) of raw materials, goods in process, and finished goods. There is an axiom—"To automate well means to understand and plan exceedingly well." Some authorities have observed that just analyzing a plant's operations and procedures when considering further automation is very worthwhile even though only a limited amount of automation may be immediately installed. For the first time, such analysis may cause an in-depth understanding of the intricacies and interrelationships of a given production situation. A number of special techniques, most supported by excellent software for computerized analysis, have been developed in recent years. These include such concepts as *group technology* and other aspects of material resources and requirements planning (MRP). These concepts are described in a condensed fashion later in this article.
2. *Enhanced product quality*, which improves competitive position and reduces waste, and reworks. Improved competitive position naturally translates into higher volume and its usual attendant economic advantages.

   It is interesting to note that some automation has entered the factory, not necessarily by choice, but rather by the force of improved manufacturing and processing operations that far exceed the limitations of human dexterity, awareness, cognition, speed, and strength, among other factors. Some manufacturing and processing variables, such as temperature, pressure, chemical composition, flow, weight, etc., are not directly measurable by people. Human inadequacies in these areas were among the first of the "external" forces that introduced a need for automation.
3. *Upward shift of workers' role*, that is, from numerous arduous, low-skilled duties to higher-skilled supervisory and maintenance responsibilities.
4. *Reduction of personal accidents* through the assumption of accident-prone duties by automated machines and processes.

---

[1] Work, as used here, is the action or effort expended in production. Work refers to the application of machine energy, human energy (muscle and brain-power), and any other auxiliary energy used in the production of goods and/or services. Work may apply only to the manipulation of information, which occurs in data processing and office automation. Or, it may apply to the manipulation of both information and materials to produce physical goods, that is, the products of industrial manufacturing and processing.

[2] D. S. Harder, who in 1947, was a vice president of Ford Motor Company.

[3] As found in "The Myth of the Machine: Technics and Human Development," by Lewis Mumford, Harcourt, Brace & World, Inc., New York, 1966.

Some of the limitations of automation include:

1. *High cost* of designing, building, and maintaining automated equipment. This cost is finding considerable relief because of the continuing lower cost of electronic components and equipment, although some of these savings are offset by continuing inflated costs of software. Successful efforts to date and that will continue into the future in terms of standardizing equipment, communication networks, and software will also relieve cost as a barrier to automation.
2. *Vulnerability to down time* because of increasing complexity of automated equipment. This vulnerability, however, is being reduced at an accelerated rate because of improved equipment self-diagnostics, fault-tolerance techniques, and more economic approaches to designing redundancy into automatic systems.
3. *Loss of flexibility*. This was a very important restraint on automation until relatively recently. Introduction and refinement of the concept of flexible manufacturing systems (FMS) has largely negated this restraint.
4. *More management attention*. Actions in highly-automated systems occur sometimes at an almost unbelievably high rate and allow little or no time for human decisions. Currently, with state-of-the-art technology, a machine or process can be driven to make quite a lot of off-spec and scrap material before effective supervision can get into control of a runaway situation. Through the assistance of information networks, ranging from corporate to plant-wide to cells and individual machines and processes, managers can be appraised of factory floor situations on essentially a second-by-second basis. Thus, more and better management personnel is needed as a plant increases its content of automation. In the more distant future, a much greater portion of the almost instantaneously needed decision making will also be done automatically. However, assignment of this important responsibility to computers needless to say will require exceedingly careful attention and analysis by very sharp management personnel.
5. *Persistance of automation's negative image*. Surprising to many authorities has been the acceptance of automation technology by the labor force and the successful negotiation of new union contracts—even though the basic fact remains that jobs are eliminated by automation. Of course, automation also creates new and certainly higher-skilled jobs. As the public and the press and other media which serve it become better acquainted with the real nature of automation, earlier predictions of very adverse effects on the labor force will continue to be tempered. Fortunately, too, automation is frequently identified with the other aspects of so-called "high technology," contributing to a reasonably good press for automation.

Applications for Automation Technology

Nearly all human endeavors, including education, recreation, health care, national defense, communication, transportation, industrial manufacturing and processing, research and development, and business and commerce have been impacted by automation.

Office Automation. Sometimes simply referred to as computerization, office automation involves information as the input, the work in process, and as the final product. The information may be of many purposes and formats—payroll preparation, transportation reservations and scheduling, banking and security transactions, statistical and census compilations, inventory control, accounts receivable and payable, insurance risks and records, cost and price analysis, statistical quality control, electronic mail, and almost any activity that can be described as routine *paperwork*. Increased productivity per office worker is indeed a major advantage, but possibly more important is the rapidity with which information required to make business management decisions can be communicated over long distances and integrated with information from various institution and corporate entities.

Office automation has contributed in a very marked way in the furtherance of manufacturing and processing automation systems.

Manufacturing Automation. Manufacturing automation, in the long term, most likely will well exceed office automation in terms of investment. However, there will be so much blending, integration, and information exchange between the management of offices and factories that it will become increasingly difficult to determine any sharp demarcation between these two activities.

The tempo to automate production has hastened very much during the latter half of the 1980s, but what appears as intense activities now will pale in terms of investments in automation to be made during the remainder of this century. The somewhat lagging acceptance of automation on the part of the bulk of manufacturing industries is considered by many authorities as simply a "wait and see" attitude. Numerous segments of manufacturing are awaiting the experiences of the comparatively few leading users of the present time, notably the application of automation technology by the automotive and electronics industries, as examples, of current leaders in the field. These industries have been under much pressure to improve both quality and productivity from forces that are national and international. Competitive pressures have warranted unusually high investments in manufacturing research and development. It is largely these industries, for example, that have funded advanced communication links and more effective robotization, including machine vision.

Patterns of Industrial Production. Manufacturing automation has developed along two principal paths, which reflect the rather distinct natures of two kinds of products:

1. *Fluid and bulk materials process industries*, as typified by the chemical, petroleum, petrochemical, metals smelting and refining, and food processing fields, among others, which largely react, separate, combine, and otherwise process materials in a liquid, slurry, gaseous, or vaporous state. During much of the manufacturing, raw materials, materials in process, and final products are in the form of fluids or bulk solids. Except at the molecular level, these materials are not in the form of discrete, identifiable pieces. Fluids and bulk materials are handled in enclosures, such as vats, bins, and other vessels, and are transported within pipes and atop bulk belt and other types of conveyors. A major exception in a number of these fluid/bulk industries is the final product which may be a discrete can, box, tankcar, barrel, etc.

   A rather high degree of automation has existed in the fluid/bulk industries for several decades, particularly since World War II when many of the former batch processes became continuous in nature. Fluid/bulk industries traditionally have been capital rather than labor intensive. For many years and continuing into the present, the most commonly measured and controlled variables have been temperature, pressure, flow, and liquid level and, as previously mentioned, these are quantities that essentially are impossible for humans to measure accurately, if at all, without the aid of instruments. The automation of measurement and control of these variables for many years was identified as *instrumentation and automatic control*—the term automation was rarely used in this regard.

   A typical automatic control system of the type used in the chemical and process industries is detailed in article on **Control System.**
2. *Discrete-piece manufacturing industries*, as typified by the manufacture of machines and parts, assemblies and subassemblies, etc., generally have been quite labor intensive because the production variables present—dimension, position, displacement, proximity, motion and speed have been at least partially within the grasp of measurement and hence control by people. Technologically, too, it has been much more difficult to develop sensors to automatically measure and devices to automatically control, without human supervision, these manufacturing variables than, for example, the development of instrumentation for the fluid/bulk industries.

   Applications of automation that typically are found in the discrete-piece manufacturing industries are illustrated by Figs. 1 through 6.

   The progress of automation has been closely tied to the ease with which an operation may be automated. Thus, it is no surprise that automation of the fluid/bulk industries preceeded the discrete-piece industries by several decades.

   This observation is further proved by a number of discrete-piece industries that currently remain well behind the leaders in

Fig. 1. Robotic system incorporating both machine vision and end-effector pressure sensors is designed for unloading randomly positioned parts from a storage bin. (*ORS-i-bot*™, ORS Automation, Inc.)

automation. For example, still one of the most labor intensive industries is the manufacture of garments and apparel. The skills of sewing have been very difficult to transfer to a machine control system. Unlike working with rigid materials such as metals and plastics, textiles are soft, pliable, and from the standpoint of machine design they are much more difficult to manipulate. Further, the geometry of the parts of a garment and the dependence for appearance upon the nature of the seam for shape and drape are factors that do not enter in the assembly of something made from harder, more rigid materials. These kinds of difficult technical problems, coupled with an industry that is generally not accustomed to high capital expenditures, have substantially slowed the pace of automation in the garment and other like fields.

Scientific Foundation for Automation

Principal scientific and technological developments that contributed to the feasibility of automation have included:

1. *Feedback*, the fundamental principle and basic mechanism that underlies all self-regulating processes. Some experts have defined feedback as information about the output at one stage of a process or operation that is returned, that is *fed back* to an earlier stage so as to influence its action and hence to change the output per se. Ingenious self-regulating devices and machines date back many years. The flyball governor, invented in 1788 to control Watt's steam engine, exemplifies the application of feedback long before a theory for feedback and closed-loop control was put forth. One of the earliest uses of closed-loop feedback was its application to the power steering of ships, adapted decades later to the power steering for automobiles.

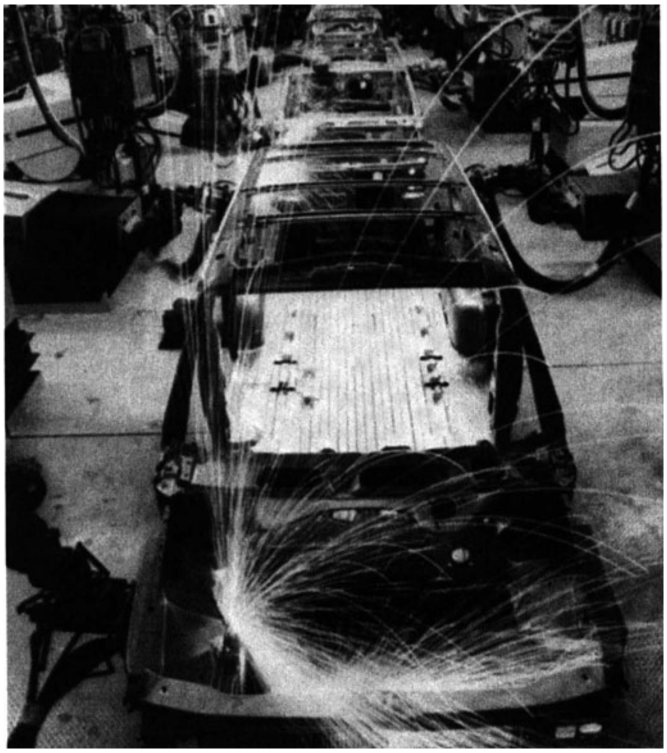


Fig. 2. Automatic welding line where the work (unitized car body parts, such as side aperture panels, roof panels, flat floor pan, and fenders) is brought to computer-controlled robots by way of conveyor line. (*Chrysler Corporation, Windsor, Ontario Assembly Plant.*)

2. *Information and communication theory* was not tackled formally until after World War II, when C. E. Shannon published "A Mathematical Theory of Communication" in 1948. In that same year, N. Wiener published "Cybernetics or Control and Communication in the Animal and the Machine." The concepts put forth by Wiener stirred up excitement during that early period. Cybernetics is essentially comprised of three concepts: (1) Animal or machine systems, (2) communication between systems, and (3) regulation or self-regulation of systems.
3. *Sensors and measurement systems* did not develop historically according to any particular master plan. Generally, sensors were developed so that more could be learned concerning the nature of physical and chemical phenomena—*not* as tools for achieving automation. Measurement of dimension and weight, for example, had its roots in antiquity and its needs were largely the basis upon which early trade could be conducted. Although mechanically based sensors have and will continue to be used in automation systems, the measurement field progressed much more rapidly after the details of electromagnetics and electrical circuits were established earlier by such investigators as Ampere, Volta, and Ohm in the late-1700s and early-1800s—then to be followed in the first half of the 1800s by Faraday, Henry, Wheatstone, Kirchoff, and Maxwell. Even before the appearance of electronics, it usually was found much easier to measure and control a machine or process by electrical rather than mechanical, pneumatic, or hydraulic means. But in the absence of electronics, nonelectrical methodologies essentially by default became the approaches of choice. Even today, wide use of mechanical, pneumatic, and hydraulic technologies persist. The comparatively new field of micromechanical sensors is successfully re-establishing some of the earlier non-electronic approaches.
4. *Servopower*, electric, hydraulic, and pneumatic, made possible a host of actuators, ranging from valves, louvers, and dampers in the fluid/bulk industries and machine and workpiece positioners in discrete-piece manufacturing. Automation was assisted by the appearance of combined-technology devices, such as electromechanical, electrohydraulic, and electropneumatic relays and subsystems. The continuing progress in the design of electric motors,

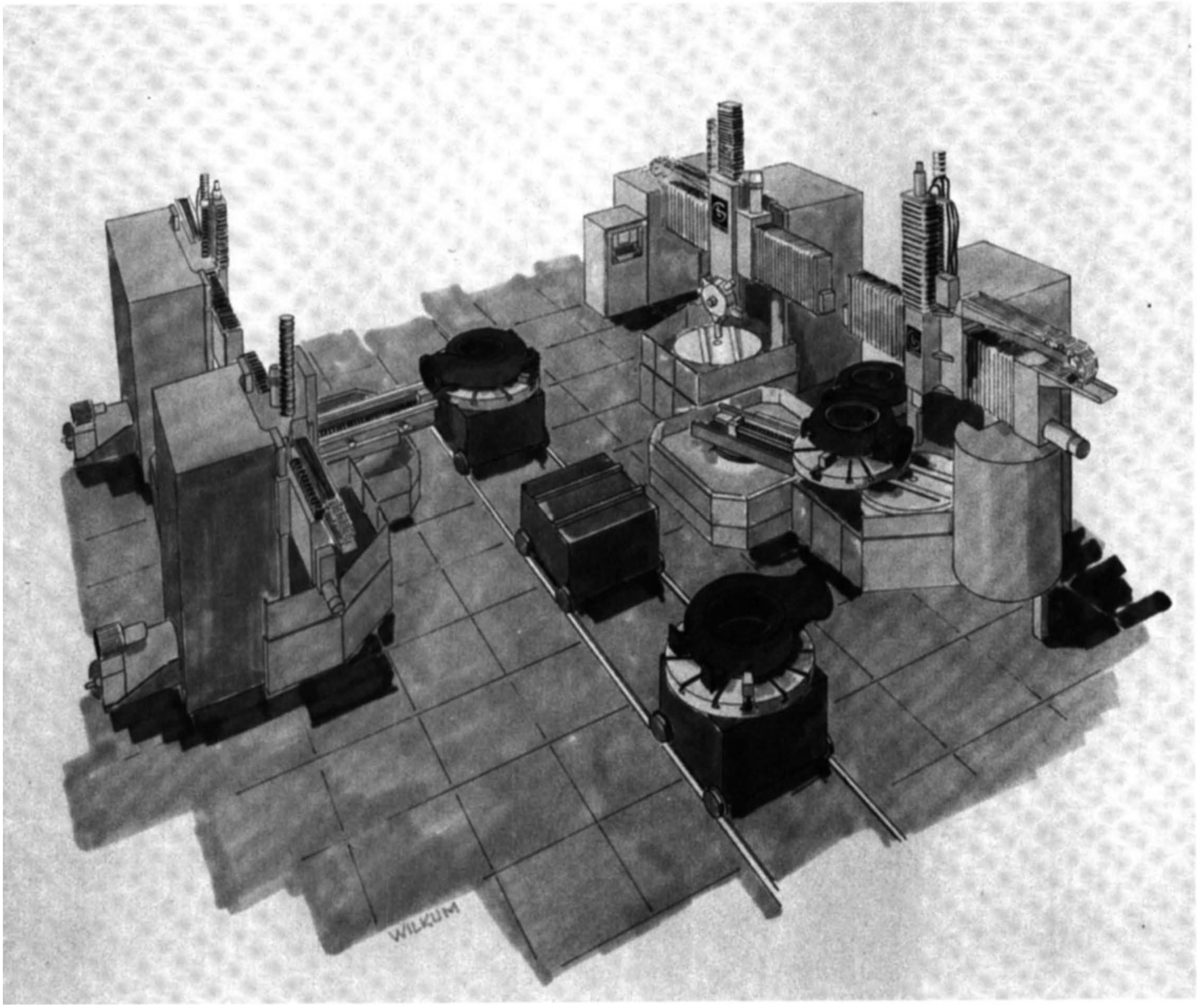


Fig. 3. Manufacturing systems can take advantage of automation even if relatively small and simple. Machining of large castings in system shown here is handled by four numerically controlled (NC) vertical turning centers with workpiece and rotary pallet automatically moved from transport to vertical turret lathe by sliding pallet shuttles. (*Giddings Lewis.*)

decreasing size and weight for a given horsepower rating along with increased energy efficiency, is contributing to the furtherance of automation. During the past decade or two, outstanding progress has been made in DC and AC motor controls, in the refinement of stepping motors, and in the practical application of linear motors.

5. *Computer and memory power* have been of outstanding importance to automation even though these elements have not always been sophisticated. The Hollerith card, which appeared in 1890 (frequently referred to for many years as the IBM card), most likely had its roots in the card-programmed Jacquard loom invented in 1801. In repeat-cycle automated machines, the memory required for operation in earlier machines was designed right into the mechanics of the machine—a practice that still can be found in printing and packaging machines, whose automaticity dates back a number of decades. As the degree of automaticity and complexity of a machine or process increases, there are continuing requirements for more information storage and retrieval at faster and faster rates. Prior to the entry of digital electronics, mechanical computing and memory systems (for example, desk calculators of just a few decades ago) were large, slow (in today's terms),

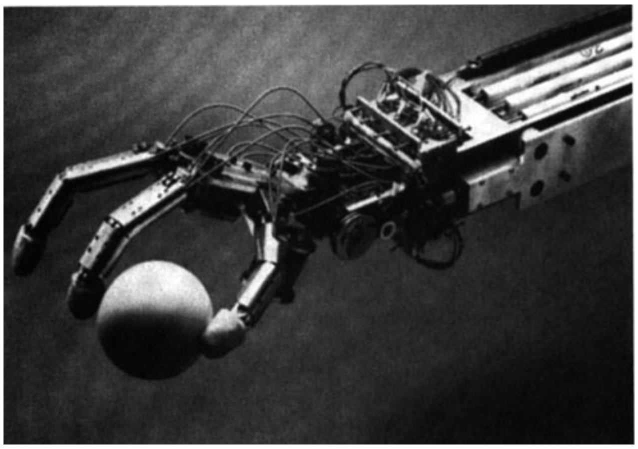


Fig. 4. Multi-finger robot hand. The hand has fourteen joints which are driven by special "shape memory" alloy actuators. The fingers are dexterous and gentle (egg in view) and are appropriate for a number of automatic assembly and maintenance operations. (*Hitachi America, Ltd.*)

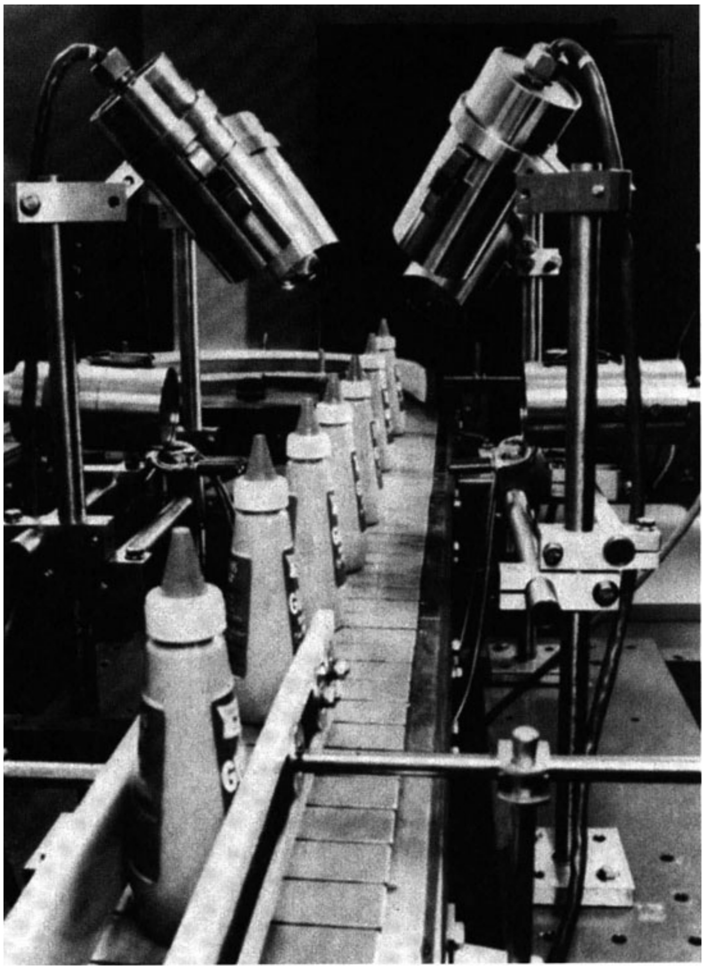


Fig. 5. Four machine-vision cameras are used in this production line to check labeling and other container characteristics. (*System QR4000, Cutler-Hammer Products, Eaton Corporation.*)

and frequently quite difficult to alter (program). With the majority of controllers of the last few decades being electronic, it is easy to forget that the earlier mechanical, pneumatic, and hydraulic controllers had to incorporate non-electronic computers to calculate the error signal in a closed-loop feedback system. Actually, the words *memory* and *computing* were rarely used in the process control field prior to the appearance of electrical and electronic instruments, even though all the elements were there under different designations.

6. *Digital technology*, which for practical purposes encompasses the advances of solid-state microelectronics, introduced vastly improved computing speeds for automated systems which, in combination with improved response speeds of detectors and sensors, greatly enhanced the performance of control systems. Modern computerization, of course, stems directly from digital technology. The two very marked trends of decreased size and cost for microelectronics have greatly influenced the availability of components in terms of application feasibility and economics. The question is sometimes asked, "Why is small size so important in regard to the electronic components widely used in automated systems?" First of all, size is directly related to the economics of component part production. Second, the example of having to mount detectors on robot arms (where the space available is limited) serves to answer the question from a practical applications standpoint. Obviously, many similar examples could be given.
7. *Mechanization*, presently simply taken for granted, was a major step toward automation. Mechanization was the logical next step toward automation after the emergence of metal hand tools (in contrast with the earlier stone and wood tools). Mechanization conferred the first degree of automaticity to a system. See article on Robots and Robotics.

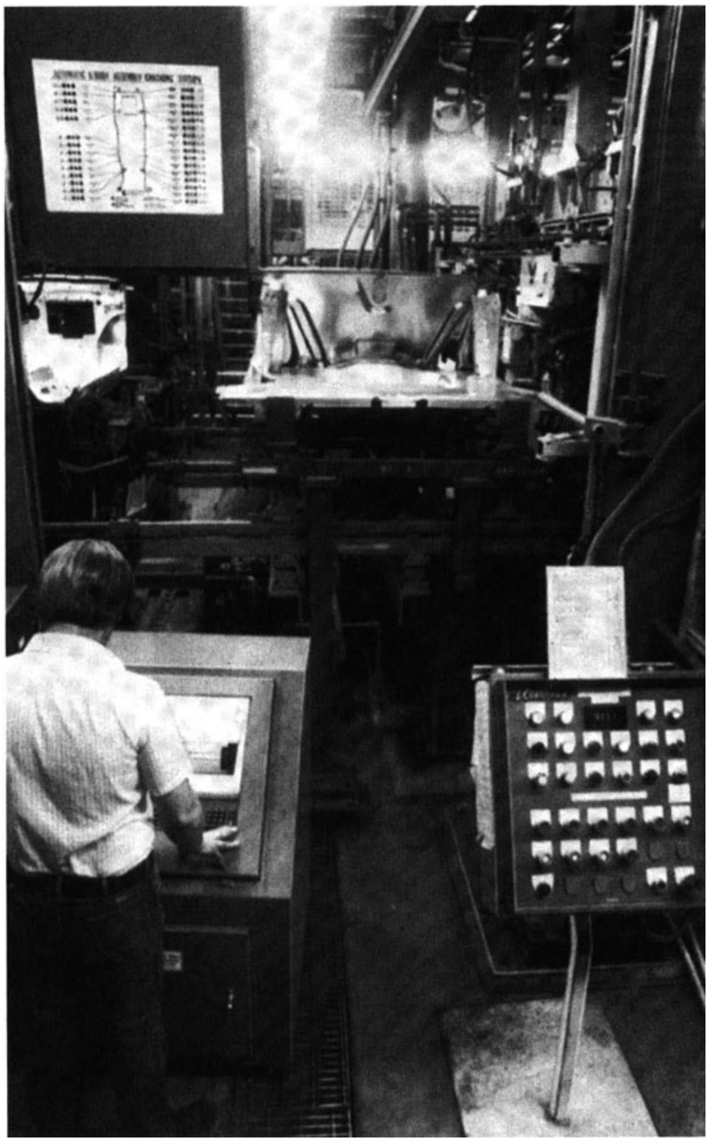


Fig. 6. After car body welding is completed, underbody subassemblies for front-wheel-drive cars move into an automatic probe checking station shown here where thirty critical measurements are verified to assure sound, dimensionally precise platforms upon which to build the rest of the vehicle. Data from the fixture are fed to a computer which prints out a copy for the operator in foreground. A board which flashes "go or no-go" lights for a quick visual verification in included. This is shown on the panel at upper left. This installation is one of over a thousand computers of varying capacity used at this Orion Township, Michigan, plant of General Motors Corporation.

8. *Systemitization and engineering analysis* were and continue to be key elements for achieving successful automation plans and installed systems. As mentioned earlier, just good planning and thinking in depth about the prospects of automation for a manufacturing process can be extremely beneficial. Traditionally, production supervisory personnel have been the real storehouse of knowledge pertaining to all aspects of production—from incoming materials through warehousing and shipping.

   Because advanced manufacturing automation minimizes (sometimes displaces) the subtleties of human judgment that can be applied directly on the factory floor in the form of minor machinery adjustments or procedural changes in the interest of maintaining smooth, uninterrupted production throughput, all of the vagaries of production which are deeply implanted in the minds of production supervisors must be brought out into the open prior to more extensive automation. This detailed, but very important information is not always easy to retrieve. As suggested by a major firm, one must "sweat out the details" if success is to be achieved via automation.

9. *Information display technology*, which has progressed beyond earlier expectations prior to the extensive use of the cathode ray tube, has contributed immeasurably to the expansion of automation technology—largely by automating the human/machine/process interface per se. Ingenious ways of plotting and presenting information, now widely assisted by the use of color, have provided a way to interlock designing for manufacture with manufacturing itself in so-called CAD/CAM (computer assisted design/manufacturing).

Contributions of Automation to Engineering and Science

The prior recitation of the scientific and engineering developments upon which modern automation is based provides only part of the story. Within the past decade, with the firm establishment of automation in many major industries, the reverse transfer of technology has occurred at least to some degree. Pressures brought about by automation have impacted information communication—as represented by the possibility (once a dream) of integrating and interlocking manufacturing operations on a corporate and plantwide basis through the development of hierarchical two-way information transfer (communication) systems. This is exemplified by the great progress that has been made in the design of local area networks (LANs), which in turn are parts of wide area networks. Many examples can be given. One of the earlier and outstanding developments is MAP (Manufacturing Automation Protocol).

The concept of distributed control is another. Introduced in the mid-1970s, this control architecture combined three technologies—microprocessors, data communications, and CRT displays. Automation today is impacting on the design of future computers, on the development of more effective programming languages, the technology of expert systems, and although not exclusively, automation is a major source of pressure to develop the concept of artificial intelligence (AI), which in past years has not exhibited the kind of practicality that is expected within the relatively near future. Progress is being made in the application of AI to machine vision in connection with the performance of robots.

Automation requirements of the automotive industry literally gave birth to the concept of the programmable controller as a replacement of electromagnetic relay systems. The acceptance of the programmable controller was almost immediate and over the past decade has expanded at a phenominal rate.

Not the least of automation's contributions to technology has been its impact on the entire philosophy of manufacturing. For example, the concept of flexible manufacturing systems (FMS). This actually grew out of earlier dissatisfaction with attempts to automate various machines and processes. With the kinds of hardware available in the 1930s and 1940s, systems were essentially limited to *hard automation*, an approach that usually was advantageous only for high-volume, long-term production runs. In fact, the popular approach to automation in the 1940s and 1950s was to design a product *for* automation (there is still wisdom in this approach). It was found that products designed strictly with automation in mind often turned out unattractive aesthetically and minimized the options in design which the consumer expected. Although no universal automated system appears on the distant horizon, automation of the late 1980s is many times more flexible. The analytical planning required to create successful flexible manufacturing systems almost immediately led to the concept of computer-integrated manufacturing (CIM). This is the logical organization of individual engineering, production, and marketing and support functions into a computer-integrated system. Functional areas, such as design, inventory control, physical distribution, cost accounting, procurement, etc. are integrated with direct materials management and shop-floor data acquisition and control. Shop-floor machines serve as data-acquisition devices as well as production machines.

Status of Automation—1994

Prior to the 1970s, the automation of industrial production was mainly an extension of mechanization, that is, the use of systems that did not incorporate feedback. Attempts to automate were largely of an unplanned, scattered, piecemeal nature. Even by the late-1980s, just a few plants worldwide (considering the vast number of manufacturing facilities, large and small) have been automated across the board in a way that matches the rather distorted public image of automation on a grandiose scale. A few notable examples will be found in the automative and electronics industries. Plantwide, *all-at-once automation* is found in a comparative handful of plants that either are new facilities built from the ground up in very recent years, or are plants that have been fully refurbished from the receiving to the shipping dock. In either case, such new and modernized facilities represent tremendous capital outlays that are well beyond the resources of most manufacturers and processors.

Plant owners and managers have patiently learned that automation is best approached by stages in a carefully planned and tightly controlled manner. What has changed most during the past few years is the attitude of top management toward automation. Greater motivation is shown because of increasing courage and confidence. Growing numbers of firms are pioneering automation on a vastly increased scale by targeting larger sections and departments of their plants—as contrasted with a former posture of experimentation and *automation by trial* in terms of a few machines or manufacturing islands.

The incentive to automate, of course, is fundamentally economic. Competitive pressures, frequently from the international marketplace, have been great and largely unexpected. Thus, any endeavor that will trim costs in the long run, such as automation, must be given serious consideration. This factor accounts for the present *uneven* application of automation from one industry to the next. Those industries that have been hurt the most by competition will be among the pioneers of automation. Very large firms in these categories not only have invested heavily in the procurement of automation hardware and software, but also have participated in a major way in automation research and development.

Technologies Closely Coupled with Automation

Closely coupled with automation are several concepts that have revolutionized manufacturing philosophy. Some of these concepts are defined briefly here and will be found in other areas of this encyclopedia. Check alphabetical index.

CAD (Computer-Aided Design)

This acronym can also be taken to mean computer-aided or computer-assisted drafting. Uncommonly, a combined acronym (CAD/D) may be used. This designates a system that assists not only in the preparation and reproduction of drawings, but that also develops the information or intelligence associated with the drawing. Most CAD/D systems have six major components (four hardware; two software):

1. A central processing unit (CPU).
2. Storage—where drawings and graphics are stored electronically.
3. Workstation—the interface between operator and computer.
4. Plotter station—where images stored in the computer memory are printed on drafting media.
5. Operating system (OS)—the master control program that coordinates the activities of all four of the aforementioned hardware components.
6. Application program—user software that creates working environment for creating designs and preparing drawings.

Major Functions of CAD. There are four principal functional categories:

1. *Design and Geometric Modeling*. In this function, the designer describes the shape of a structure with a geometric model constructed graphically on a cathode ray tube (CRT). The computer converts picture into a mathematical model, which is stored in the computer database for later use.
2. *Engineering Analysis*. After creation of a geometric model, the engineer can calculate such factors as weight, volume, surface area, moment of inertia, center of gravity, among several other characteristics of a part. One of the most powerful methods for analyzing a structure is *finite element analysis*. Here, the structure

is broken down into a network of simple elements and the computer uses these to determine stress, deflections, and other structural characteristics. The designer can see how a structure will behave before it is built and can modify it without building costly physical models and prototypes. The procedure can be expanded to a complete systems model and operation of a product can be simulated. When combined with engineering, CAD is sometimes referred to as CAE (Computer-Aided Engineering); or sometimes the combined acronym, CAD/CAE, is used.

3. *Computer Kinetics*. The user can examine effects of moving parts on other parts of the structure or design and analyze more complex mechanisms.
4. *Drafting*. A CAD system can automatically draft drawings for use in manufacturing. Engineers can draw on geometric and numerically coded descriptions produced by CAD to create numerical control tapes, which permit direct computer control of shop machines, determine process plans and scheduling, instruct robots, computerize testing, and generally improve the management of plant operations.

CAM (Computer-Aided Manufacturing)

This acronym generally refers to the utilization of computer technology in the management, control, and operation of a manufacturing facility through the direct or indirect interface between a computer and the physical and human resources found in a manufacturing organization. Developments in CAM are found in four main areas:

1. *Machine Automation*. Originally confined to numerical control, machine automation has been expanded and now consists of a chain of increasingly sophisticated control techniques:
   (a) At the lower end of the scale is *fixed automation* with relays or cams or timing belts and timing chains. Relay logic has been extant in industrial production for decades. Essentially during the past two decades, many relay installations have been replaced by electronic means, notably in the form of programmable controllers.
   (b) Further up the scale of automaticity is plain numerical control (NC) whereby a machine is controlled from a pre-recorded, numerically coded program for fabricating a part. In these systems, machines were hard-wired and were not readily reprogrammable.
   (c) At a higher point in the scale of automaticity, the machine is directly controlled by a minicomputer, which stores machining instructions as software that is relatively easy to reprogram. Known as CNC (computer numerical control), this approach has the advantages of much higher storage capability and greatly increased flexibility. Nearly all new numerical control systems today are CNC oriented. However, as recently as the late 1970s, CNC was considered a costly exception to the traditional approach.
   (d) At the highest point in the scale of automaticity as presently viewed is the plant and even corporate-wide interconnection of machines on the floor with vast and complex information networks wherein decisions at the factory floor level are influenced by information flowing down from the corporate computer hierarchy—and, in the other direction, information from machines flows upward to enrich the database of the headquarters computer. This is further described under CIM.
2. *Robotics*. Robots are now used rather widely for performing materials-handling and manipulating functions in CAM systems. Robots can select and position tools and workpieces for CNC tools, operate such tools as drills and welders, or perform test and inspection functions.
3. *Process Planning*. This activity considers the detailed sequence of production steps from start to finish. The process plan describes the state of the workpiece at each work station. An important element of process planning is *group technology*, in which similar parts are organized into families to allow standardized fabrication steps, thus permitting savings by avoiding duplicate tooling and system engineering. This approach differs some from traditional practice where parts were usually fabricated close to their assembly into a subsystem. Similarity of fabrication techniques may take precedence over what otherwise might appear as the logical location for such work.
4. *Factory Management*. This involves the coordination of operations of an entire plant and is a key objective of CIM. Systems tie together individual machine tools, test stations, robots, and materials handling systems into *manufacturing cells* and the cells are integrated into a larger system. Integrated management requires extensive, highly detailed, and usually costly software programs. The nomenclature used for this software is not consistent. The term Manufacturing Planning and Control Systems (MPCS) has been used as a grand designation. Two very important areas in this grouping are MRP-I (Materials Requirement Planning) and MRP-II (Manufacturing Resources Planning).

Bridging CAD and CAM Systems

CAD employs pictorial, graphics-oriented computer databases, whereas CAM involves a large amount of text-oriented information. It is necessary to find a way for the computer doing the drawing to speak the same language as the computer directing the manufacturing plant.

Layering is one way to link the systems. This involves structuring the CAD and CAM databases. This makes it possible for various people to input data without losing control of the overall design and manufacturing process. Also, it permits shop people to see information that is meaningful to them without sorting through and attempting to understand the rest of the information that is normally included in a drawing. This can be accomplished by organizing information into an arrangement resembling layers or slices within the databases. The engineers or users in other departments of an organization can provide pertinent information or examine any or all layers of information in accordance with need.

CIM (Computer-Integrated Manufacturing)

The concept of CIM was introduced in 1973. It has been defined by some authorities as the automation and integration of a manufacturing enterprise through the use of computers. Others have defined CIM as the logical organization of individual engineering, production, and marketing/support functions into a computer-integrated system. Functional areas, such as design, inventory control, physical distribution, cost accounting, planning, purchasing, etc., are integrated with direct materials management and shop-floor data acquisition and control. Thus, the loop is closed between the shop floor and its controlling activities. Shop-floor machines serve as data-acquisition devices for the control system and often its direct command. Strategic plans smoothly give way to tactical operations, at known cost.

CIM is obviously a very ambitious target. The internal research into nearly every aspect of a firm's business is required to prepare complex software for CIM. This is very time consuming and costly. Until there is some universality of approach and until it becomes easier to follow the successful CIM examples of others, the average plant may need a number of years on the learning process prior to adopting CIM on a large scale. At present, CIM appears to be most practical for very large firms, particularly those with multiple product lines that change every year or so (notably, the automotive industry).

Fortunately, many of the gains from automation can be made without having to go to the high technological level demanded by CIM.

FMS (Flexible Manufacturing System)

The definition of FMS has widened over the last few years. Traditionally, FMS has been considered to be the mingling of numerical control (NC) with automated materials handling and computer systems. When simple tape-programmed NC was introduced in the late 1950s, a major selling point was the ability of such a system to permanently remember from data stored on punched paper tape how given parts were made. This enabled manufacturers to run batches of the same parts with long time intervals between batches. All that was necessary was to retrieve the tape from file and to run the machine tool on a fresh supply of blanks. Duplication from batch to batch was quite exacting. Finished part inventories could be trimmed drastically. Better quality control over parts and workmanship were key objectives of the early NC systems. The ability to store tapes was an added benefit of the system and proved to be one of the main advantages as experience was gained.

Improvements in computer control and the creative concept of automated material transport between machines spawned what is known as a flexible manufacturing system. Now, instead of applying CNC (computer numerical control) to only one or at most two machine tools, groups of machine tools can be operated from the same controls. Direct labor still may be used to unload and load workpieces in a centralized area and to handle tool replacements. These functions also can be fully automated if economically justified. Generally, each individual palletized workpiece is automatically sent to required work stations in the unique order appropriate for its processing. The order of processing and the actual work done may differ for each part in the system at any one time. Provisions are made for automatic rerouting of parts if a given workstation becomes unavailable or overloaded. The computer system notifies the system manager of any malfunctions, monitors tool life, and signals any required normal tool replacements. Production, down time, and individual machine utilization statistics are recorded and reported as required.

Because FMS installations are among the most common applications for automation found in industry, they are mentioned frequently in the literature. These systems may be extended in complexity and sophistication to whatever extent management may be willing to fund. The FMS concept is not limited to machine tools.

Motion Control Systems

A majority of equipment that fits the definition of automation does not involve robots. However, robots were an important incentive toward the development of motion control. The control of motion is a key function of automated manufacturing systems, whether or not robots per se are involved. Dating back a few decades, motion sensors of high precision and reliability were developed, many of which incorporate stepping motors as a source of motion power. See also article on **Stepper Motors.** Other machine drives, of course, include gear drives, tangential drives, and lead or ball screw drives. See Figures 7 through 11.

D. M. Considine, P.E., and Glenn D. Considine

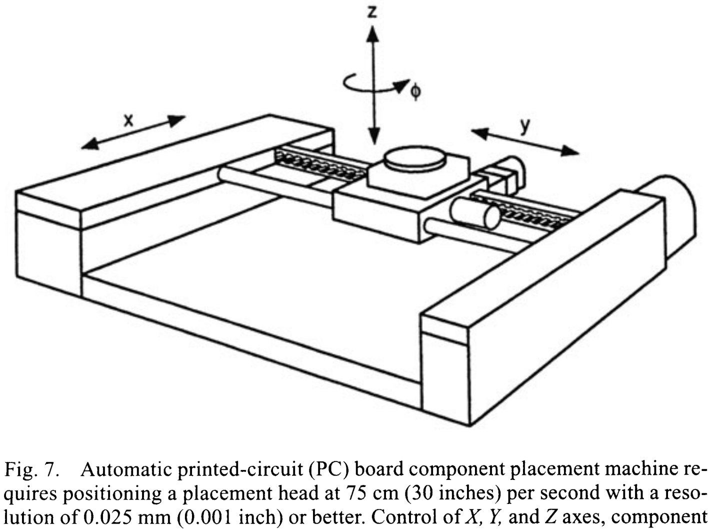


Fig. 7. Automatic printed-circuit (PC) board component placement machine requires positioning a placement head at 75 cm (30 inches) per second with a resolution of 0.025 mm (0.001 inch) or better. Control of X, Y, and Z axes, component alignment, and gripper are required from computer programming. A belt-driven gantry is controlled by an indexer, and two servomotor drives are used for X-Y positioning. Z motion and rotational alignment are controlled by a computer microstepping drive. Joystick inputs are used to move the head manually and to teach positions to the computer. (*Parker-Hannifin Corporation, Compumotor Division.*)

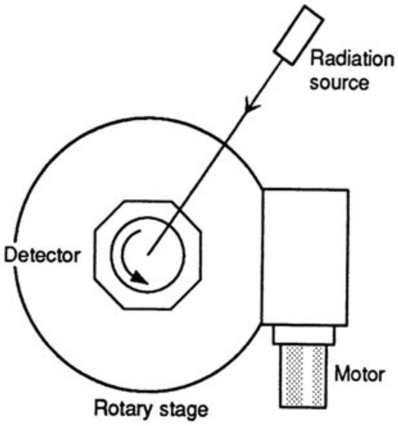


Fig. 8. A system is required to plot the response of a sensitive detector, which must receive equally from all directions. Detector is mounted on a rotary table which requires to be indexed in 3.6° steps, completing each index within 1 second. (For setting up purposes, the table can be positioned manually at 5 r/min. The table incorporates a 90:1 worm drive.) The maximum required shaft speed (450 r/min) is well within the capacity of a stepper, which is an ideal choice in simple indexing applications. Operating at a motor resolution of 400 steps per revolution, the resolution at the table is a convenient 36,000 steps per revolution. In this case it is important that electrical noise be minimized to avoid interference with the detector. Two possible solutions are to use a low-EMI linear drive or to shut down the drive after each index. (With a stepper driving a 90:1 worm gear there is no risk of position loss during shutdown periods.) (*Parker-Hannifin Corporation, Compumotor Division.*)

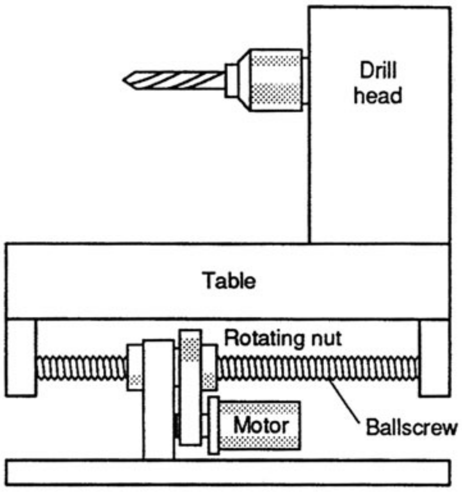


Fig. 9. A stage of a transfer machine is required to drill a number of holes in a casting using a multihead drill. The motor has to drive the drill head at high speed to within 2.5 mm (0.1 inch) of the workpiece and then proceed at cutting speed to the required depth. Drill is now withdrawn at an intermediate speed until clear of the work and then fast retracted, ready for the next cycle. Complete drilling cycle takes 2.2 seconds, with a 0.6-second delay before the next cycle. Due to proximity of other equipment, the length in the direction of travel is very restricted. An additional requirement is to monitor the machine for drill wear and breakage. The combined requirements of high speed, high duty cycle and of monitoring the drill wear all point to use of a servomotor. By checking torque load on the motor (achieved by monitoring drive current), one can watch for increased load during the drilling phase, pointing to a broken drill. Application will require a ball-screw drive to achieve high stiffness together with high speed. One way of minimizing the length of the mechanism is to attach the ball screw to the moving stage and then rotate the nut, allowing the motor to be buried underneath the table. Since access for maintenance will then be difficult, a brushless motor is suggested. (*Parker-Hannifin Corporation, Compumotor Division.*)

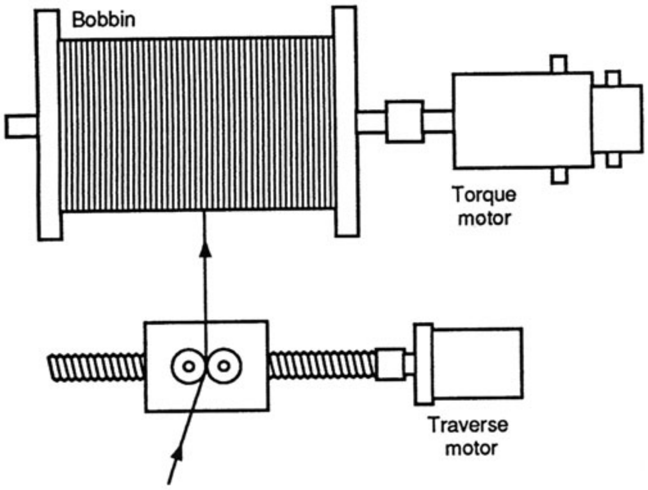


Fig. 10. Monofilament nylon is made by an extrusion process which results in an output of filament at a constant rate. Product is wound onto a bobbin rotating at a maximum speed of 2000 r/min. Tension in filament must be held between 0.1 and 0.3 kg (0.2 and 0.6 lb) to avoid stretching. Winding diameter varies from 5 to 10 cm (2 to 4 inches). Prime requirement is to provide a controlled tension, which means operating in a torque mode rather than a velocity mode. If the motor produces a constant torque, the tension in the filament will be inversely proportional to the winding diameter. Since the winding diameter varies by 2:1, the tension will fall by 50 percent from start to finish. A 3:1 variation in tension is acceptable, so constant-torque operation is acceptable. Requirement leads to use of a servo operating in the torque mode. (Need for constant-speed operation at 2000 r/min also makes a stepper unsuitable.) Rapid acceleration is not needed, so a brush servo would be adequate. In practice, this suggests a servo in velocity mode, but with an overriding torque limit. The programmed velocity would be a little over 2000 r/min. In this way the servo will normally operate as a constant-torque drive, but if the filament breaks, the velocity would be limited to a programmed value. The traversing arm can be adequately driven by a stepper. However, the required speed will be very close to resonance, so a microstepping system would be preferable. An alternative would be to use a half-step drive in conjunction with a toothed-belt reduction of about 3:1. A ball-screw drive can be used to achieve high stiffness together with high speed. (*Parker-Hannifin Corporation, Compumotor Division.*)

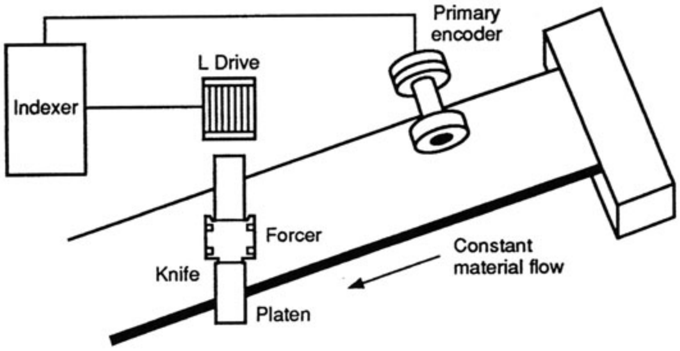


Fig. 11. Plastic sheet cutting. Process produces a continuous flow of sheeted plastic to be cut into prescribed lengths before it is fully cured. Material is cut as it exits a machine and cannot be stopped. Depending on ambient conditions, the speed can vary. Clean angle cuts are required. In system shown, an encoder is mounted to a friction wheel driven by the plastic material. This speed signal is an input to a self-contained indexer (controller), which references all linear (cutoff-knife) velocity and position commands to the encoder, allowing precise synchronization of the web. Placing the knife at an angle to the flowing material allows for precise, straight cuts while material is moving. This is an excellent application for a linear motor. (*Parker-Hannifin Corporation, Compumotor Division.*)

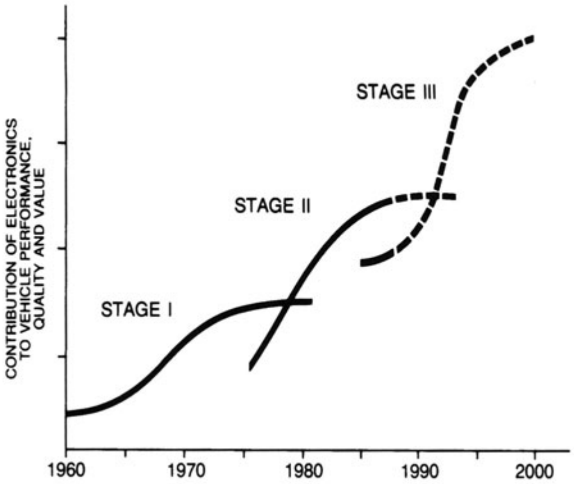


Fig. 1. Three states in the evolution of automotive electronics from the early beginnings in the 1960s to the contemplated vehicles of the year 2000. (*Ford Motor Company, Electronics Division.*)

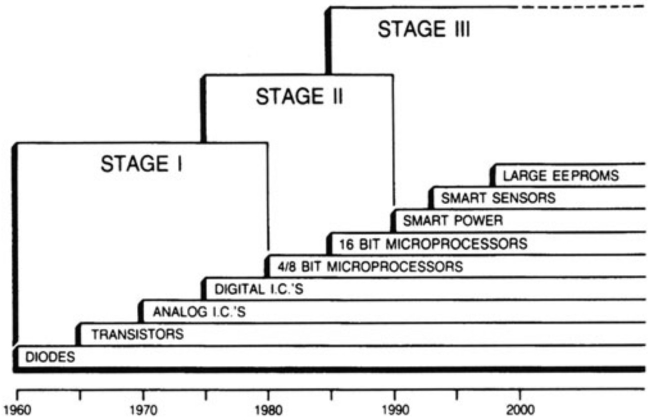


Fig. 2. Semiconductor evolution corresponding with the three stages in the evolution of automotive electronics. (*Ford Motor Company, Electronics Division.*)

AUTOMATISM. See **Brain Injury; Coma.**

AUTOMOTIVE ELECTRONICS.[1] Electronic devices and systems over the last several years have changed the character of the automotive vehicle in numerous ways, including engine control, the suspension system, steering, power train, braking, comfort, safety, and entertainment systems, in addition to fuel economy, among other important factors. As this article points out, this trend is destined to continue through the year 2000 and for the unforeseeable future.

Although solid-state electronics has been applied to motorcars and other automotive vehicles since the 1960s, the use of electronics in their design and operation did not start its impressive rise in terms of availability and consumer acceptance until the mid-1970s. As noted in Fig. 1, the steady and rapid rise commenced in the 1980s and has continued through 1993. The extent to which electronic systems have been integrated into motor vehicles has depended largely on the availability of new and high-performing electronic components and systems at an acceptable cost, coupled with much research on the part of the automotive manufacturer to determine how electronics can cost effectively improve vehicle performance, safety, and comfort, without adding excessive costs. It will be noted that one manufacturer (Ford), for historical and

---

[1] This is one of a series of articles in this *Encyclopedia* that discuss various aspects of automotive technology, including construction materials and alternate vehicle fuels and energy sources. See *Alphabetical Index*.

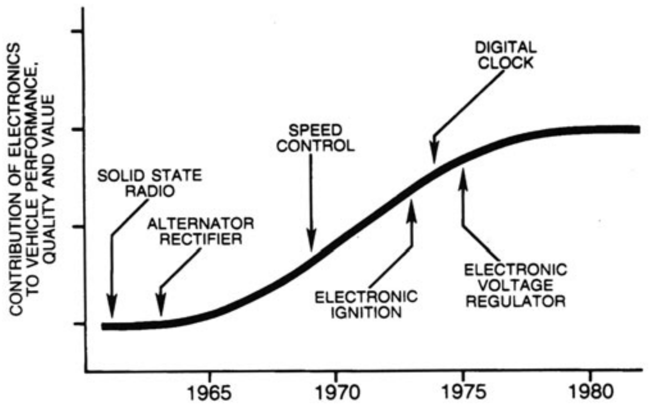


Fig. 3. Use of stand-alone electronic components typical of the early applications in automotive electronics. (*Ford Motor Company, Electronics Division.*)

planning purposes, has broken down the trends of Fig. 1 into three phases or stages. These are further depicted in Fig. 2.

Stage One—1960s to Mid-1970s. Solid state electronic devices were first widely used in automobiles in the 1960s. Over the following 15 years, diodes, discrete transistors, and analog integrated circuits (ICs) were used to solve problems in stand-alone electronic components. See Fig. 3. The earliest installation in autos (Ford) of an all solid-state radio occurred in 1961. Solid-state ignitions were first incorporated in 1973. See Fig. 4. A stand-alone electronic clock appeared in the instrument panel in 1974. These early applications provided a learning experience with the emerging electronic technology. More importantly, they demonstrated that electronics could achieve the levels of reliability required by the automotive industry, while providing dependable and affordable service to the customer. For example, new electronic devices increased the dependability of many automobile components. Radio failures became much less frequent. Car clocks became more reliable. Electronic ignition eliminated the routine maintenance associated with changing the breaker points.

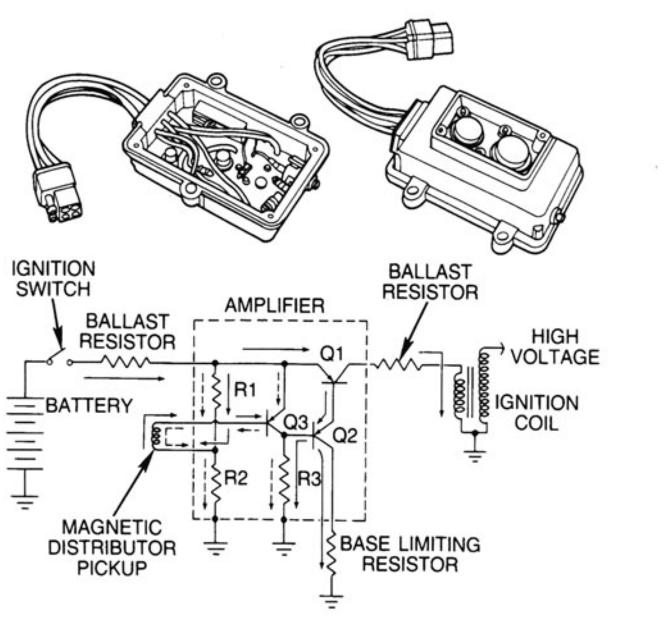


Fig. 4. Solid-state ignition module and circuit diagram. First used in 1973. (*Ford Motor Company, Electronics Division.*)

Stage Two—Mid-1970s to Early 1990s. In the mid-1970s, new electronic capabilities became available to the automotive industry. Their applications were sufficiently different to define another stage in the evolution of automotive electronics.

The primary source of new capability was the microprocessor. It came along at a time when the industry was experiencing upheaval. Gas lines, safety standards, and emission controls created unprecedented challenges. The microprocessor led to many solutions.

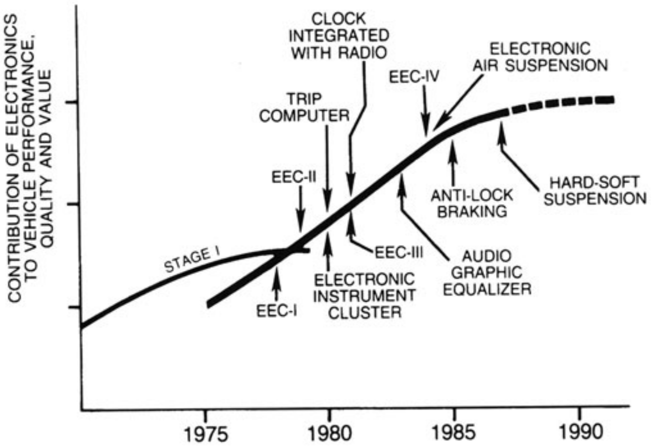


Fig. 5. Linking of electronic components into fewer, more sophisticated systems typifies automotive electronic technology during the period from the late 1970s to the late 1980s. (*Ford Motor Company, Electronics Division.*)

Stage Two was characterized by a shift from independent components to increasingly sophisticated systems which link components together. See Fig. 5. These systems first were used for engine controls. For example, Ford introduced the EEC-1 in 1978. Several sensors were linked with a computer which, in turn, was linked to various output devices, such as the ignition module. Similarly, multiple electrical and electronic components were tied together in driver information and entertainment applications. Electronics have been used to integrate functions which previously, for example, had been separate clock and radio displays.

Late Stage Two developments will make more widespread use of advanced packaging techniques, such as surface-mounted devices, increased processor speed and capacity, increased memory, and refined input/output (I/O) methods to obtain optimal functional performance from the auto's special-purpose subsystems.

Powertrain control will be adapted to a greater range of conditions, and there will be more integration between engine and transmission. Anti-lock braking will be more common. See Fig. 6. Improved displays will offer drivers more choices of format and content. Cellular telephone communications will be increasingly common. All of these improvements should create vehicles that are more responsive, useful, and reliable.

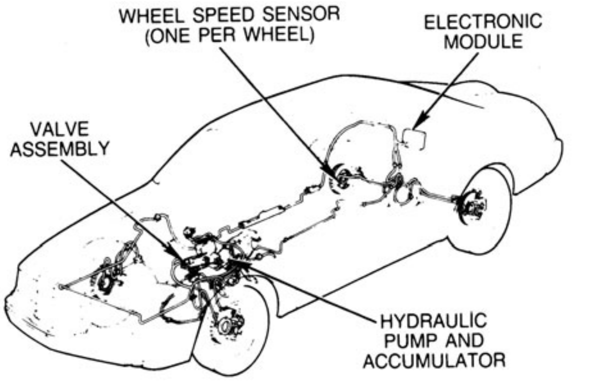


Fig. 6. Four-wheel anti-lock braking system. This is one of the first features of Stage Three in the evolution of automotive electronics. (*Ford Motor Company, Electronics Division.*)

As the vehicle's subsystems evolve during the mid-to-late 1990s, they will form a total vehicle *network*. Sensors, processors, and actuators will be interconnected, with power and control signals distributed in a highly efficient manner. This functional integration will lead to Stage Three.

Stage Three—1990 to 2000+. This development phase of automotive electronics will be characterized by the emergence of a vehicle with a totally integrated electrical and electronic system. Designers will escape from the mechanical function replacement and "add-on" approaches that characterized Stages One and Two. They will seek to optimize the performance of the total vehicle through electronics. The total system will have much greater flexibility and adaptability, with extensive software control of multifunction features. There will be greater opportunities for auto buyers to customize their vehicles. Vehicle characteristics, such as ride quality, handling properties, steering effort feedback, and brake "feel" information will be incorporated in a display format, and even engine power versus economy trade-offs will be selectable and controllable by the driver.

Operating as an information-based system, the automobile's on-board electronics will use extensive computing capacity, multiplexed circuit technology (networks), and program memory capacity that will be considered very large by present (1992) standards.

Some early Stage Three features are beginning to appear. See Fig. 7. These include speed control integrated with engine control and transmission controls integrated with engine controls. See Fig. 8. These examples represent only the beginnings in the systematic integration of functions. Stage Three will introduce: (1) torque-demand powertrain control, which will fully integrate the response of the engine and transmission; (2) vehicle dynamics, which will integrate braking, steering, and suspension; (3) electric power management based on new power generating components and sophisticated load management controls, and (4) multipurpose soft switches and shared displays for driver information, climate control, and entertainment functions.

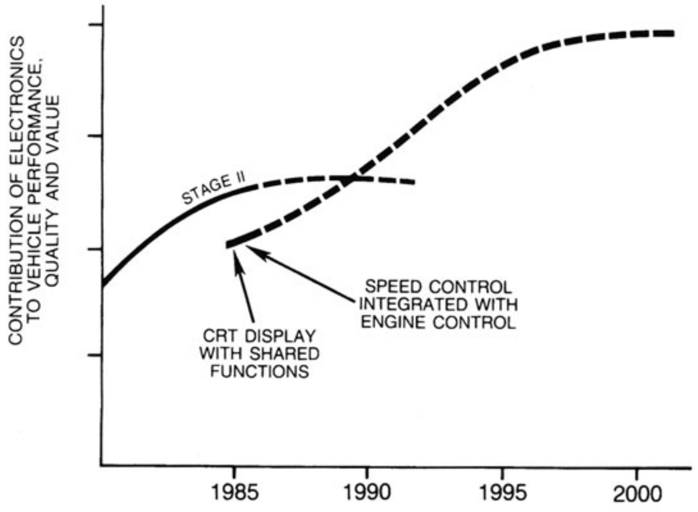


Fig. 7. Stage Three will evolve around functional integration of electronics. (*Ford Motor Company, Electronics Division.*)

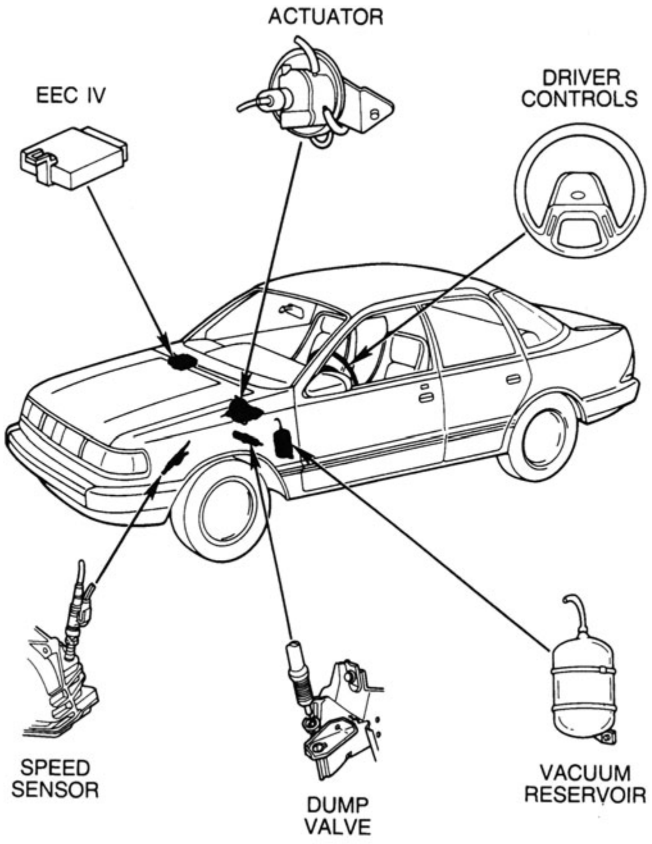


Fig. 8. Integration of speed control system electronics into electronic engine control is another example of the fast-moving trend in electronics that will lead to the automobile of the year 2000. (*Ford Motor Company, Electronics Division.*)

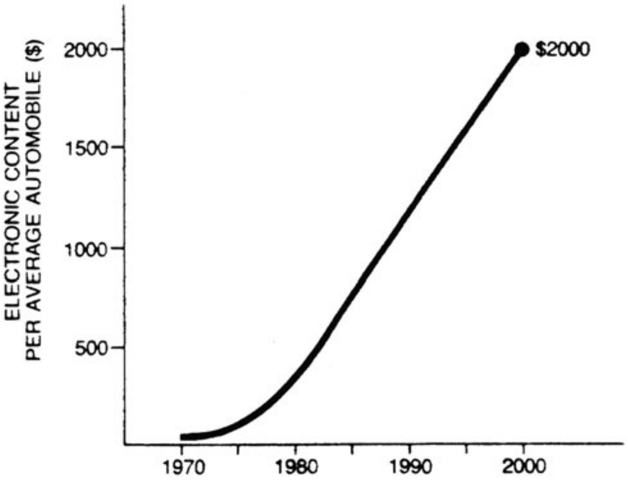


Fig. 9. Automotive electronics content for an average car will reach $2,000 by the year 2000. (*Ford Motor Company, Electronics Division.*)

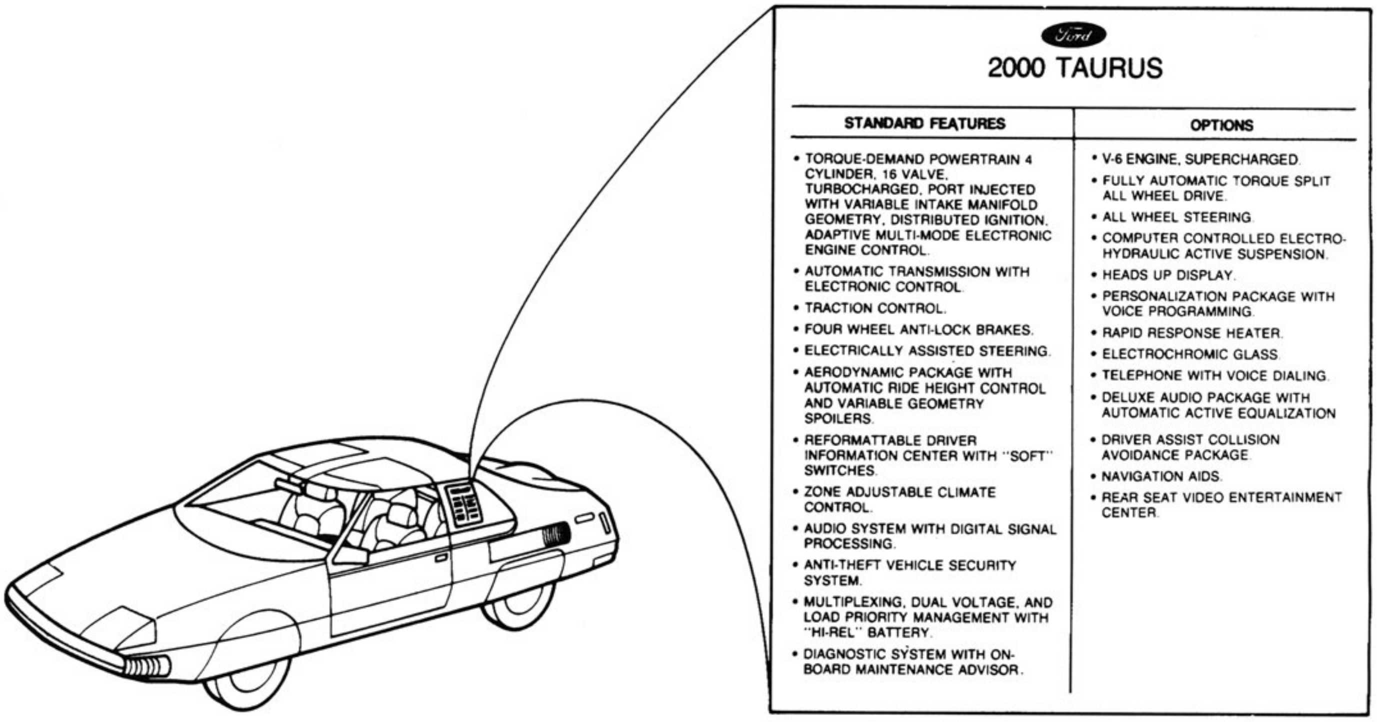
Stage Three, now underway, is planned to be fully evolved by the year 2000. At that time it is predicted that the value of electronics for the average automobile will reach $2,000. See Fig. 9.

The realization of Stage Three will incorporate impressive new capabilities, as reflected by what may be a typical window sticker for a car in the year 2000. (In predicting some of these features, the assumption is made that fuel availability and cost will be similar to the 1990's time frame. Should fuel become scarce or very costly, a different picture could emerge.) See Fig. 10.

Powertrain. The engine compartment will contain a lightweight, supercharged or turbocharged four-cylinder or six-cylinder, multi-valve engine of 1.5 to 2.5 liters displacement. See Fig. 11. It will be equipped with multi-point electronic fuel injection and distributed ignition, which is *distributorless*, and will have a high-voltage coil at each spark plug. The system will control the engine on a cylinder-by-cylinder basis.

The powertrain will be electronically controlled by a highly advanced system. Engine operating parameters will be adaptively controlled over the full range of torque and RPM, and will allow the driver to select for either performance or economy. Variables under active and continuous electronic control will include manifold boost pressure, fuel mixture, spark timing, valve timing, and variable intake manifold geometry.

The information needed to manage the engine control will come from a small number of high-performance sensors. The sensors will monitor, analyze, and transmit data on fundamental engine performance parameters. Improved sensors will be necessary before this is possible. Primary data will include combustion chamber conditions and exhaust gas chemistry. These data will be compared by the master controller to a



Ford

2000 TAURUS

STANDARD FEATURES

- TORQUE-DEMAND POWERTRAIN 4 CYLINDER, 16 VALVE, TURBOCHARGED, PORT INJECTED WITH VARIABLE INTAKE MANIFOLD GEOMETRY, DISTRIBUTED IGNITION, ADAPTIVE MULTI-MODE ELECTRONIC ENGINE CONTROL.
- AUTOMATIC TRANSMISSION WITH ELECTRONIC CONTROL.
- TRACTION CONTROL.
- FOUR WHEEL ANTI-LOCK BRAKES.
- ELECTRICALLY ASSISTED STEERING.
- AERODYNAMIC PACKAGE WITH AUTOMATIC RIDE HEIGHT CONTROL AND VARIABLE GEOMETRY SPOILERS.
- REFORMATTABLE DRIVER INFORMATION CENTER WITH "SOFT" SWITCHES.
- ZONE ADJUSTABLE CLIMATE CONTROL.
- AUDIO SYSTEM WITH DIGITAL SIGNAL PROCESSING.
- ANTI-THEFT VEHICLE SECURITY SYSTEM.
- MULTIPLEXING, DUAL VOLTAGE, AND LOAD PRIORITY MANAGEMENT WITH "HI-REL" BATTERY.
- DIAGNOSTIC SYSTEM WITH ON-BOARD MAINTENANCE ADVISOR.

OPTIONS

- V-6 ENGINE, SUPERCHARGED.
- FULLY AUTOMATIC TORQUE SPLIT ALL WHEEL DRIVE.
- ALL WHEEL STEERING.
- COMPUTER CONTROLLED ELECTRO-HYDRAULIC ACTIVE SUSPENSION.
- HEADS UP DISPLAY.
- PERSONALIZATION PACKAGE WITH VOICE PROGRAMMING.
- RAPID RESPONSE HEATER.
- ELECTROCHROMIC GLASS.
- TELEPHONE WITH VOICE DIALING.
- DELUXE AUDIO PACKAGE WITH AUTOMATIC ACTIVE EQUALIZATION
- DRIVER ASSIST COLLISION AVOIDANCE PACKAGE.
- NAVIGATION AIDS.
- REAR SEAT VIDEO ENTERTAINMENT CENTER.

Fig. 10. Envisioned "window sticker" on a year 2000 Taurus with Stage Three electronics. (*Ford Motor Company, Electronics Division.*)

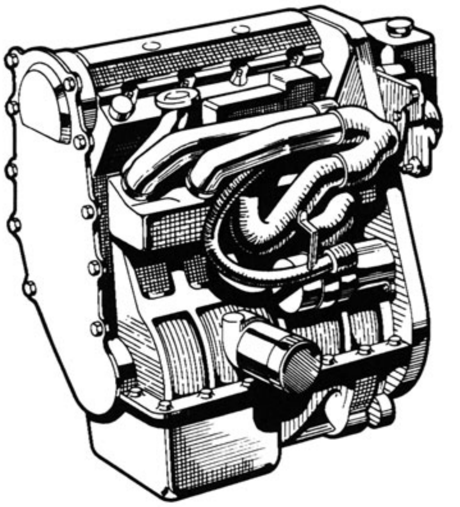


Fig. 11. Four-cylinder engine performance will improve with Stage Three electronics. (*Ford Motor Company, Electronics Division.*)

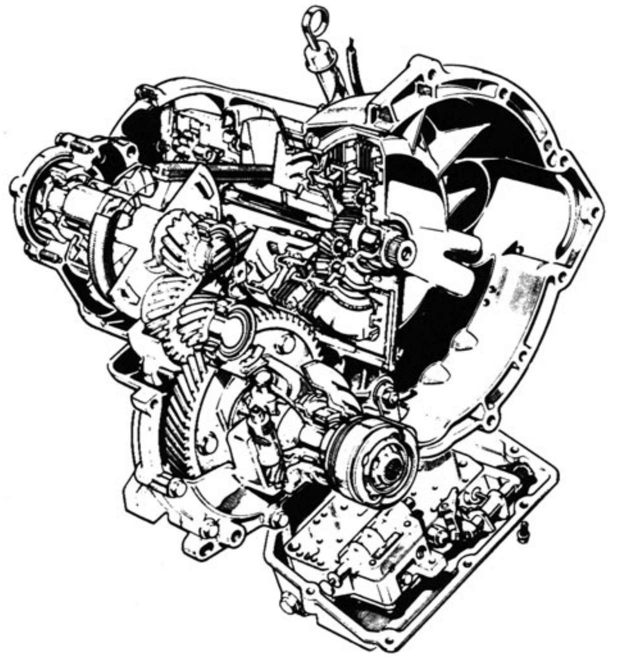


Fig. 12. Continuously variable transmission (CVT). (*Ford Motor Company, Electronics Division.*)

performance algorithm of much greater sophistication than exists today. Then all controlled variables will be adjusted to optimize vehicle performance in accordance with the algorithm. This will be a significant improvement over current systems which measure secondary parameters, such as inlet charge temperature and barometric pressure.

The engine will be closely coupled with a transmission of advanced design. The transmission either will be continuously variable (see Fig. 12) or will have a highly adaptive shifting capability. Transmission and engine will be electronically controlled as a unit in response to the driver's demand for power. Responding to engine speed, vehicle speed, and command input from the driver, the powertrain controller will decide when to supply torque by increasing engine output, altering the drive ratio, or both. Figure 13 contrasts the degree of powertrain integration typical of Stages Two and Three. In some cases, the transmission will drive all four wheels and utilize a sophisticated electronic control scheme.

It is planned that drivers in the year 2000 will find that the powertrain will perform smoothly under all conditions. Changes in transmission ratio and adjustments in engine speed will seem nearly imperceptible as compared with those in present cars. The powertrain management system also will enhance safety and improve performance in ways that are meaningful to all users (i.e., by keeping the driver in control at all times).

Chassis. Perhaps the most revolutionary new features and performance enhancements resulting from Stage Three systems integration will occur in the area of the chassis system, consisting of steering, brakes, and suspension. Drivers will experience a new level of performance from these chassis systems through their synergistic interaction, achieved through an integrated electronic network, as shown in Fig. 14.

Electronics capability will allow adaptive control of the suspen-

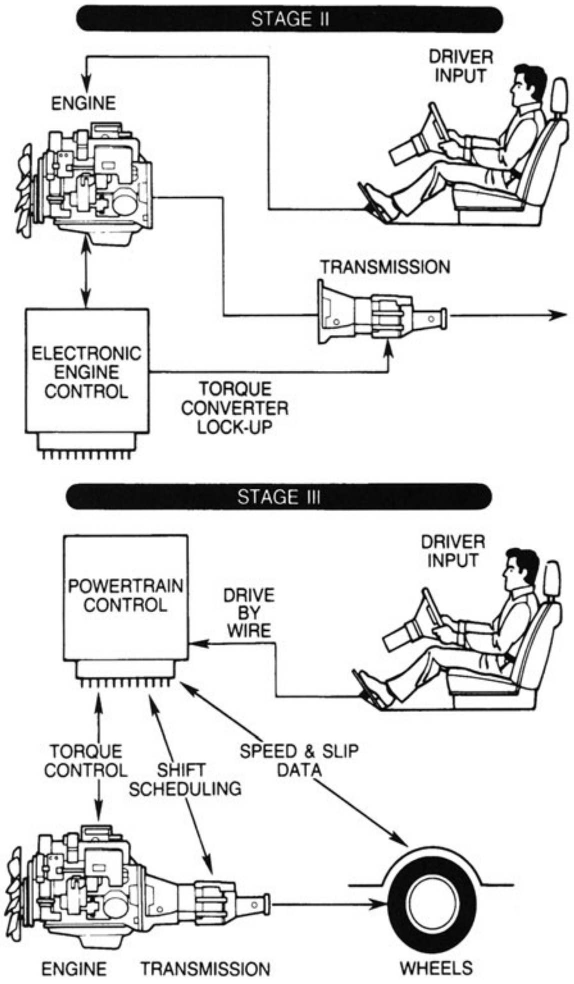


Fig. 13. Electronic powertrain systems for Stages Two and Three. (*Ford Motor Company, Electronics Division.*)

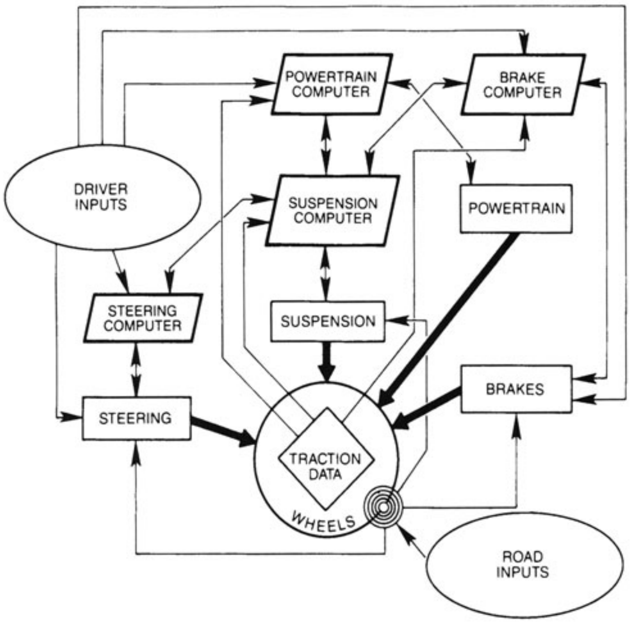


Fig. 14. Stage Three chassis control system. (*Ford Motor Company, Electronics Division.*)

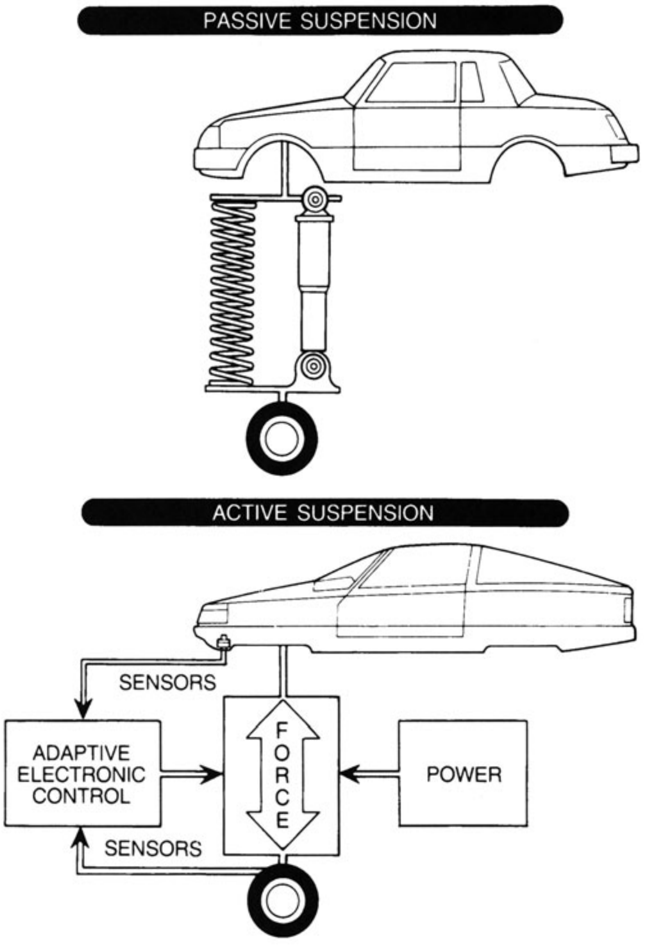


Fig. 15. Suspension evolution to active control. (*Ford Motor Company, Electronics Division.*)

sion—springs, shock absorbers, and suspension geometry—which were restricted to a passive response in Stages One and Two. See Fig. 15. In order to control ride height, aerodynamic angle of attack, and dynamic response of the body, an electronic system will sense displacements and accelerations in the suspension system and will control spring rate and damping independently, at each wheel. A first step in this direction was the introduction of ride height control in 1984. See Fig. 16.

In the year 2000, electrically actuated front-wheel steering will be found on most model cars. In the advanced cars, the suspension control will be *semi-active*. There will be continuous modulation of devices like valves to control shock absorber damping, but with no external power input. In some high-performance vehicle applications, however,

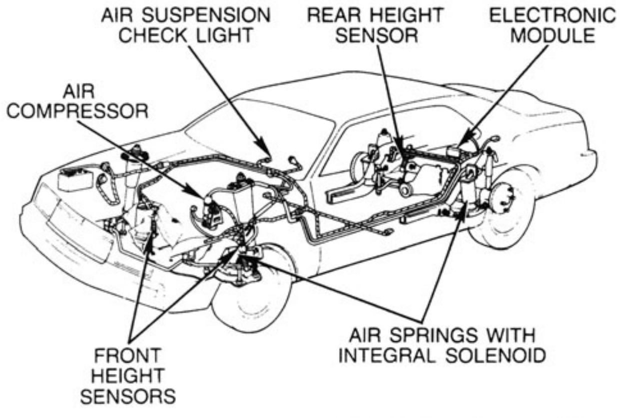


Fig. 16. Air suspension ride-height control. (*Ford Motor Company, Electronics Division.*)

fully active suspensions will incorporate controlled energy input from a dedicated power source. While the ultimate evolution of both semi-active and active systems will probably be electromechanical, those systems in use by the year 2000 will use electrohydraulic control units. In either case, the driver will experience a marked improvement in both handling quality and ride comfort.

Electrically actuated front-wheel steering (Fig. 17) will be found on most cars by the year 2000. Among the advantages of these systems will be compactness, energy efficiency, and adaptability. Also, most cars will incorporate all-wheel steering. See Fig. 18. This will improve agility at all speeds and will enhance low-speed maneuverability. Control of the steering angle of the rear wheels will be electronic, changing according to both the vehicle's speed and input from the steering wheel. Drivers of these cars will find that parallel parking is simplified and maneuvering in tight quarters takes little effort. On the highway these changes will be accommodated promptly.

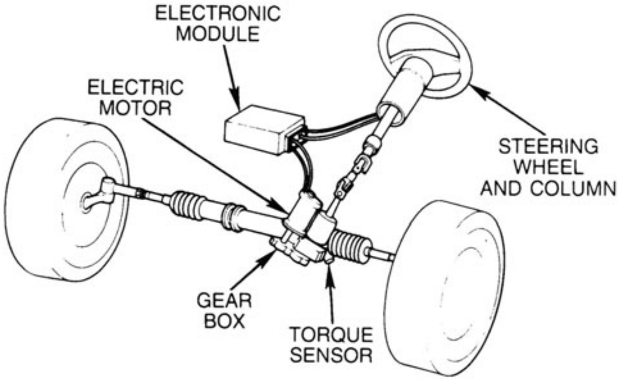


Fig. 17. Electric power steering. (*Ford Motor Company, Electronics Division.*)

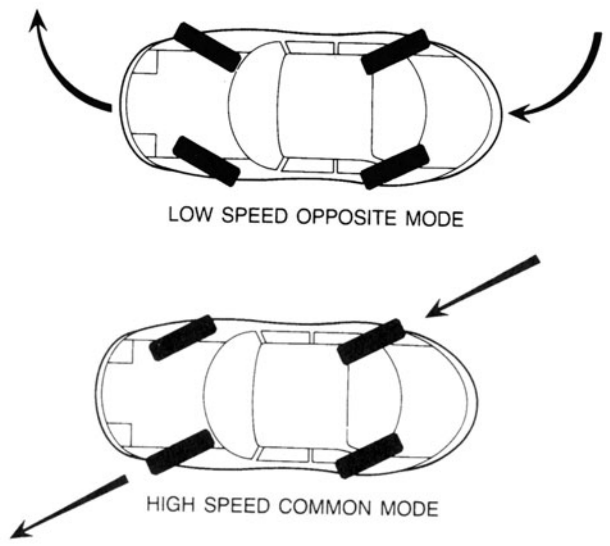


Fig. 18. All-wheel steering. (*Ford Motor Company, Electronics Division.*)

Braking performance will improve steadily as it evolves from the anti-lock braking systems that presently are becoming widely available. By the year 2000 traction control systems will fully integrate braking with the powertrain. The functional flexibility of these systems will be far greater than the simple "anti-lock" capability. Conditions which could cause slipping will be monitored during both acceleration and deceleration. The system will modulate torque and braking inputs to provide maximum acceleration and minimum stopping distance. The driver will be unable to break the wheels loose from the road under any normal driving conditions.

An enhanced traction control system will improve the vehicle's ability to avoid collisions. In some applications the space all around the vehicle will be monitored for the presence of collision risks, using some combination of sensing technologies, such as radar, laser, visual, infrared, or ultrasonic. Not only will the area in front of the vehicle be scanned to detect a rapidly closing interval, but the "blind spots" on the rear quarters will be monitored to assure safe lane changes. See Fig. 19. The output of the sensors will be analyzed by artificial intelligence (AI) software that will direct controllers to reduce acceleration or, in extreme cases, to apply the brakes and tighten seat belts.

For a period of time up to the present, the handling capability of an automobile has been beyond the average driver's skills. Advances in steering, braking, and suspension technology in Stage Three will allow the average driver to employ the full performance potential of the vehicle in exceptional situations (avoiding accidents), without subsequent loss of control. The subtle and rapid corrections needed to deal with the complex dynamic transients will be handled automatically.

Driver Information/Personalization. By the end of Stage Three, human factors design for driver information displays and controls will be markedly advanced. "Cockpit workload" will be reduced. Essential information, such as vehicle speed, will be provided continuously by a holographic heads-up display. See Fig. 20. This will be similar to what is presently used in some aircraft. Other information will be displayed on a reformattable multifunction display panel. See Fig. 21. It will use one or several display technologies, such as liquid crystals, vacuum flourescent devices, or light-emitting diodes (LEDs).

The system will display performance data whenever it senses something unusual. Also, the driver will be able to select a particular array of information and the style in which it is presented. For example, the driver will be able to request a complete display of all engine operating

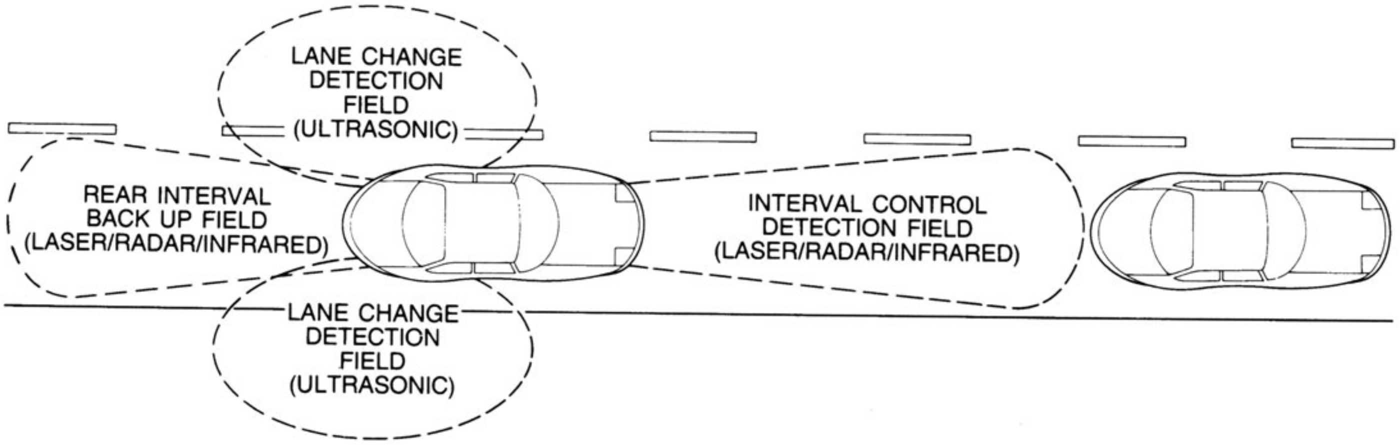


Fig. 19. Collision avoidance aids require sensors to monitor the space around the car. (*Ford Motor Company, Electronics Division.*)

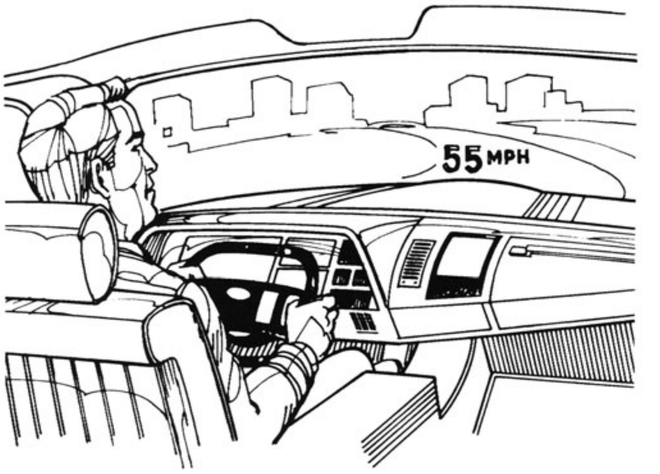


Fig. 20. "Heads-up" display of vehicle speed. (*Ford Motor Company, Electronics Division.*)

Fig. 22. Voice recognition for dialing cellular telephone. (*Ford Motor Company, Electronics Division.*)

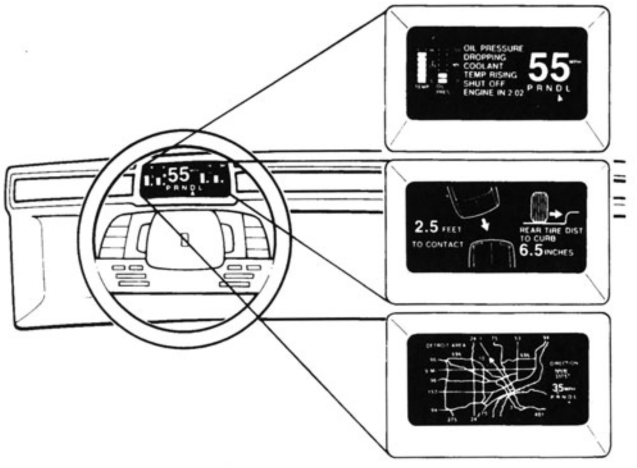


Fig. 21. Reformattable driver information center. (*Ford Motor Company, Electronics Division.*)

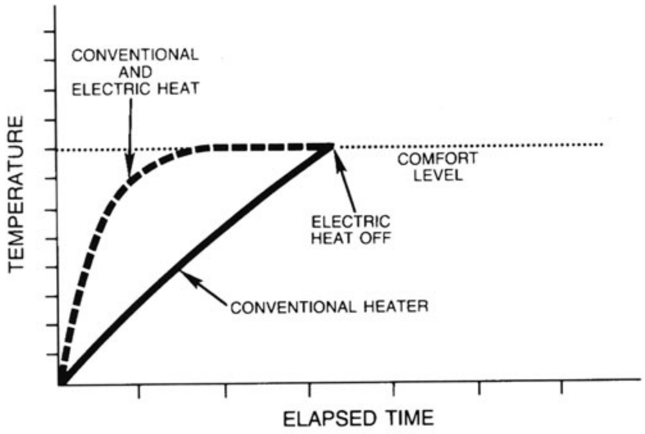


Fig. 23. Auto heater achieves comfort level faster with addition of electric heat. (*Ford Motor Company, Electronics Division.*)

parameters, such as RPM, oil pressure, coolant temperature, fuel pressure, and so on. Or the operator will be able to select data in an analog or digital format. Additional information that can be called up will include maintenance information, such as need for lubrication, brake checking, and so on.

Notification of emergency and alarm conditions will be either audible or visible signals. Voice recognition (see Fig. 22) will be used for functions, such as entertainment system control, driver display mode selections, and telephone dialing. Additional control inputs will be made by using programmable multifunction "soft" switches.

Systems also will include detection of impaired or loss of alertness on the part of the driver. They will focus on actions that are fundamental to the driver's safe operation of the vehicle, such as appropriate steering and braking behavior.

Climate Control. This system will be electronically controlled and electrically powered. The air distribution system will be designed to permit unique temperature variations in different zones of the passenger compartment. Supplemental electric heat will be capable of fast response to peak heating demand (see Fig. 23) and will allow for a reduction of excess heater capacity.

Communication and Navigation. In addition to cellular telephones previously mentioned, navigation aids will be especially useful in commercial delivery, service, and rental vehicles. These systems will operate on the combined principles of dead-reckoning and map matching, utilizing a wide-area digital map that will be stored on a compact disk read-only memory. By the year 2000, global positioning satellites may be used to track a vehicle's geographic location. These systems also will utilize an external data link that will enhance the "on-vehicle" navigation system by providing the driver with current traffic information. There also will be real-time satellite updates on roads and new landmarks.

Summary

Although this article has concentrated on the high points between the present and the goals for the year 2000 and has divided progress in automotive electronics into three stages, obviously there will be more to come during the next several decades. Progress will depend upon further developments in the electronics industry, including "smart" sensors, "smart" actuators, and advances in communications technology, among many other factors that contribute to automotive engineering.

Technical information for this article was furnished
by the scientists and engineers of Ford Motor Company,
Electronics Division, Dearborn, Michigan,
and is gratefully acknowledged by the Editors.

AUTONOMIC NERVOUS SYSTEM. The term *autonomic* signifies automatic or unconscious activity. Thus, the autonomic nervous system is sometimes referred to as the involuntary nervous system; rarely, the vegetative nervous system. A functional division of the nervous system would consist of ganglia, nerves, and plexuses, through which visceral organs, heart, blood vessels, glands, and smooth muscle receive their innervation. It is widely distributed over the body, especially in the head and neck, and in the thoracic and abdominal cavities. The autonomic system is not under voluntary control and the processes in which it is concerned are beneath consciousness for the most part. It is influenced to a great degree by the endocrine glands, particularly the adrenal and its hormone, epinephrine.

In general the autonomic nervous system may be divided into two groups both of which may send nerves to the same organs but act antagonistically, producing opposite results. One is known as the parasympathetic, which arises from the mid-brain, hind-brain, and sacral region of the cord and is stimulated by the drug pilocarpine and inhibited by atropine. The other is known as the sympathetic, which arises from the thoracic and lumbar regions of the spinal cord and is stimulated by epinephrine.

Under normal conditions there is a balance between the two systems allowing for perfect function of a bodily organ. For instance, the heart is slowed by the parasympathetic system and accelerated by the sympathetic. Movement of the stomach is increased by the parasympathetic and is inhibited by the sympathetic. The pupil of the eye is contracted by the parasympathetic and dilated by sympathetic stimuli.

Psychosomatic disturbances may take place in any of the involuntary organs of the body systems. These include the digestive, the respiratory, the heart and circulatory, the genitourinary, the endocrine system, and the skin. Gastrointestinal reactions, such as nervous diarrhea, will affect various individuals with different degrees of intensity. The emotional component of diarrhea has been recognized for centuries. It also has been recognized for centuries that certain skin disorders contain an emotional element. Eruptions arising from emotional disturbances are termed *psychogenic* skin eruptions. Most persons who show the characteristics of a nervous dermatitis, like other psychosomatic patients, appear to carry their emotional problems close to the surface of their minds, rendering them more accessible to psychiatric treatment.

As pointed out by Nauta and Freitag (*Sci. Amer.*, **241**, 3, 109, 1979), "The autonomic nervous system is not self-governing at all. Its functions are integrated with voluntary movements no less than with motivations and affects. In short, its roots are in the brain; one's experiences from moment to moment dictate not only the contractions of one's skeletal muscles but also large functional shifts in the body's internal organs. The term autonomic has nonetheless won out in the English-speaking world. Other languages use other terms. In German one speaks of *das viszerale Nervensystem*, in French of *le système nerveux végétatif*."

See also **Nervous System and The Brain.**

AUTOREGRESSION. A stochastic relation connecting the value of a variable at time t with values of the same variable at previous times. For example the linear equation

$$u_t = \alpha_1 u_{t-1} + \alpha_2 u_{t-2} + \epsilon_t \tag{1}$$

where ϵ_t is a random variable. Two common forms of autoregressive relations are the Markov scheme

$$u_t = \alpha_1 u_{t-1} + \epsilon_t \tag{2}$$

and the Yule scheme (1).

These equations bear a formal resemblance to the equations of linear regression—hence the name—but raise special problems in the estimation of the constants. They may be regarded as a class of stochastic processes.

See also **Stochastic Process.**

AUTOTOMY. Self-mutilation. Through the presence of a special modification near the base of the limb, some crustaceans and insects are able to drop off appendages by which they are seized. The autotomy of the arms of starfish and of the tails of lizards are other common examples. Autotomy is followed by regeneration.

AUTOZOOID. Members of polyp colonies whose function is to feed the colony.

AUTUNITE. This mineral is a hydrous phosphate of calcium and uranium, crystallizing in the tetragonal system, usually in thin tabular crystals. Good basal cleavage; hardness, 2–2.5; specific gravity, 3.1; luster, subadamantine to pearly on the base; color, lemon yellow; streak, yellow; transparent to translucent; strongly fluorescent.

Originally from near Autun in France, whence the name, it is a secondary mineral associated commonly with uraninite. In the United States, it occurs sparsely in the pegmatites of Connecticut, New Hampshire and North Carolina. Autunite also is known as *calco-uranite.*

See also **Uraninite.**

AUXOMETER (or Auxiometer). An apparatus for measuring the magnifying power of a lens or any optical system.

AUXOSPORE. An auxospore is a special type of spore which occurs in diatoms and which seems to be a means of rejuvenating the cells. Rejuvenescence is necessary, since in the normal process of cell division one of the two daughter cells is always smaller than the parent cell. Consequently very small cells are ultimately formed. In some species of diatoms, auxospore formation is preceded by the escape of the protoplast from the walls of the cell. The free protoplast then enlarges and secretes about itself a wall. In time new valves more or less like those of the original diatom are formed. In other species of diatoms, auxospore formation is preceded by the union of the protoplasts of two similar diatom cells, the process being therefore sexual.

AVALANCHE (Electronics). The term avalanche is used in counter technology to describe the process which is essentially a cascade multiplication of ions. In this process, an ion produces another ion by collision, and the new and original ions produce still others by further collisions, resulting finally in an "avalanche" of ions (or electrons). The terms "cumulative ionization" and "cascade" are also used to describe this process.

The term avalanche or avalanche effect is sometimes applied to the Zener effect in semiconductors.

AVALANCHE (Geology). A large mass of snow, ice, soil, or rock, or mixtures of these materials, falling or sliding very rapidly under the force of gravity. Velocities may sometimes exceed 500 km/hour. Avalanches can be classified by their content, such as snow and ice avalanches, debris avalanches, soil or rock avalanches. (*Glossary of Geology*, American Geologic Institute.)

AVERAGE. A simple but subtle concept which attempts, in some sense, to summarize a set of numbers $x_1, \ldots, x_n$ in a single number. In statistics, the commonest forms of average are

(a) The arithmetic mean M, defined by

$$M = \frac{1}{n} \sum_{j=1}^{n} x_j$$

(b) The geometric mean G, defined by

$$\log G = \frac{1}{n} \sum_{j=1}^{n} \log x_j$$

(c) The harmonic mean H, defined by

$$\frac{1}{H} = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{x_j}$$

When the individual numbers x are not regarded as of equal importance, they may be weighted by numbers $w_1, \ldots, w_n$. For example the weighted arithmetic mean is given by

$$\frac{1}{\sum_{j=1}^{n} w_j} \sum_{j=1}^{n} (w_j x_j)$$

See also **Arithmetic Mean**.

AVERAGE DEVIATION. If $\bar{x}$ is the mean of observations $x_1, \ldots, x_n$, the mean deviation is given by

$$M.D. = \frac{1}{n} \sum_{j=1}^{n} |x_j - \bar{x}|$$

If x has a frequency distribution $f(x)$ the analogous definition is

$$M.D. = \int_a^b f(x)|x - m|\, dx$$

where m is the mean and the distribution ranges from a to b.

Owing to its relative mathematical intractability the mean deviation is usually discarded in favor of the standard deviation.

The average deviation is a minimum when deviations are measured from the median.

AVOGADRO CONSTANT. The number of molecules contained in one mole or gram-molecular weight of a substance. The most recent value is $6.0220943 \times 10^{23} \pm 6.3 \times 10^{17}$. In measurements made by scientists at the National Bureau of Standards (Gaithersburg, Maryland) and announced in late 1974, the uncertainty (as compared with previous determinations) of the number was reduced by a factor of 30.

AVOGADRO LAW. The well-recognized principle known by this name was originally a hypothesis suggested by the Italian physicist Avogadro, in 1811, to explain the puzzling rule of proportional volumes observed in chemical reactions of gases and vapors. It states simply that equal volumes of all gases and vapors at the same temperature and pressure contain the same number of molecules. Though this assumption accords with the facts and aids the kinetic theory of gases, just why it should be true is by no means self-evident, unless one starts with the much more recent Maxwell-Boltzmann law of equipartition of energy, which also requires proof. That Avogadro's law is true cannot be said to have been positively established until the experiments of J. J. Thomson, Millikan, Rutherford, and others determined the value of the electron as an electric charge and thereby made it possible to count the number of atoms of different elements in a gram. The actual number of molecules contained in one mole (gram-molecular weight) of a substance is the Avogadro constant.

At any fixed temperature and pressure, the density of carbon dioxide gas, for example, is approximately 22 times greater than the density of hydrogen gas. Thus, the mass of 1 liter of carbon dioxide is 22 times the mass of 1 liter of hydrogen gas. According to Avogadro's principle, the number of molecules in 1 liter of carbon dioxide is the same as the number of molecules in 1 liter of hydrogen. Thus, it follows that a carbon dioxide molecule must have a mass that is 22 times larger than the mass of a hydrogen molecule. Since the molecular weight of hydrogen (H_2) was set equal to 2, carbon dioxide was assigned a molecular weight of 22×2, or 44. Cannizzaro was the first to use gas densities to assign atomic and molecular weights. Avogadro's principle also may be used to assign molecular weights in a slightly different way. At standard temperature and pressure, the volume of a mole of any gas is 22.4 liters. The molecular weight of a gas, therefore, is the mass (in grams) of 22.4 liters of the gas under standard conditions. For most gases, the deviation from this ideal value is less than 1%. See also **Avogadro Constant;** and **Combustion.**

AVULSION. A sudden cutting off or separation of land by a flood, or by an abrupt change in the course of a stream, as by a stream breaking through a meander or by a sudden change in current whereby the stream deserts its old channel for a new one. Generally, in legal interpretation, the part thus cut off or separated belongs to the original owner.

AXES (Aircraft). Three fixed lines of reference, usually centroidal and mutually perpendicular. The horizontal axis in the plane of symmetry, usually parallel to the thrust axis, is called the longitudinal axis; the axis perpendicular to this in the plane of symmetry is called the normal axis; and the third axis perpendicular to the other two is called the lateral axis. Rotation may take place about any or all axes; translation may take place along any of the three axes. The important translational axis is the longitudinal. See also **Bank (Aircraft); Pitching Moment; Yaw (Aircraft).**

AXIAL MAGNIFICATION. The ratio of the interval between two adjacent image points on the axis of an optical instrument to the interval between the conjugate object points.

AXIAL ORGAN. An organ of peculiar structure and unknown function found near the axis of the body in all echinoderms except the sea cucumbers.

AXIL. The angle between the upper side of a leaf and the stem to which the leaf is attached. See **Stem (Plant).**

AXINITE. This mineral is an aluminum-boron-calcium silicate with iron and manganese, $(Ca, Mn, Fe)_3Al_2BSi_4O_{15}(OH)$. It crystallizes in the triclinic system, yielding broad sharp-edged forms, which has led to its name, derived from the Greek word meaning axe. It breaks with a conchoidal fracture; hardness, 6.5–7; specific gravity, 3.22–3.31; luster, vitreous; colors, brown, blue, yellow and gray; transparent to translucent.

Axinite occurs in granites or more basic rocks along contacts and in cavities in Saxony, Switzerland, France, England, Tasmania, and Japan; in the United States, in New Jersey, Pennsylvania, and California.

AXIOM. A statement of an abstract notion which is assumed without proof. Axioms constitute the unproved first principles which are used in founding a mathematical discipline. Physical science disciplines sometimes are developed using the axiomatic approach which embraces the presentation of a minimum number of statements (axioms) and derives the other relationships common in the discipline, using only these axioms and mathematical and logical processes.

AXIS. A line so situated that various parts of an object are symmetrically located in relation to it. Also the line passing through the origin of a coordinate system which corresponds to all points of a given variable when other variables are zero. Thus, in two dimensions, the X-axis is the locus of all points whose Y-coordinate is zero. See also **Ellipse; Hyperbola; Mineralogy; Parabola.**

AXIS (Instantaneous). In rigid body motion, a line perpendicular to the plane of the motion which passes through that point or those points of a body which are instantaneously at rest. For a cylinder rolling down an inclined plane without slipping, the instantaneous axis is the line of contact between cylinder and plane.

AXIS OF ROTATION (Fixed). The locus of points of a system along a straight line which remain stationary when the system undergoes motion of rotation.

AXIS (Optic). A direction through a doubly-refracting crystal along which no double refraction occurs. A uniaxial crystal has one such direction, a biaxial has two such directions. See **Crystal.**

AXIS (Optical). The line through the foci and the vertices of the optical surfaces. Commonly, the surfaces of lenses and mirrors are figures of revolution about the optical axis. Normally, the parts of an optical system are all coaxial.

AXOLOTL (*Amphibia, Urodela*). A salamander, *Ambystoma tigrinum*, found near Mexico City which, although related to some of the terrestrial salamanders, retains its larval form throughout life, becoming sexually mature in this stage. Under experimental conditions the animal has been caused to undergo the usual metamorphosis.

AXON. The impulse-transmitting part of a nerve cell or neuron. The other parts are the cell body, containing the nucleus, and the dendrites, branches which pick up impulses. The axon is also known as the nerve fiber. The axon of a peripheral nerve of a vertebrate animal is typically covered by an inner myelin sheath and by a thin outer cellular layer called the *neurilemma*. In certain invertebrates, such as the squid, giant axons are found which range in size from 150–700 micrometers in diameter. This unusually large size has made possible many fundamental measurements on the biophysical properties of the excitable cell membrane of the axon. See also **Nervous System and The Brain.**

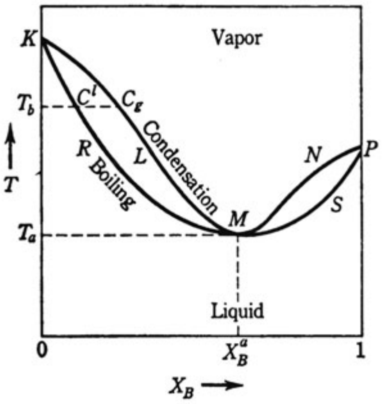
In contrast to the continuous excitory level of the neuron as a whole, the outflow of energy through an axon exhibits an *all* or *none* response. Thus, it is either quiescent or carries one or more discrete intermittent nerve impulses—only their frequencies differ.

The axon terminals of a neuron may be few in number and concentrate their effect on one or only a small localized group of post-synaptic cells. In other situations, collateral branches arise from an axon at points throughout its course and each of these may itself arborize and diverge to many destinations. The "many-to-one" convergence of information channels with the integration of their effects and, in other situations, the "one-to-many" divergence of channels are essential features of all advanced forms of nervous systems.

R. C. V.

AZEOTROPIC SYSTEM. A system of two or more components which has a constant boiling point at a particular composition. If the constant boiling point is a minimum, the system is said to exhibit *negative azeotropy*, if it is a maximum, *positive azeotropy.*

Consider a mixture of water and alcohol in the presence of the vapor. This system of two phases and two components is divariant (see **Phase Rule**). Now choose some fixed pressure and study the composition of the system at equilibrium as a function of temperature. The experimental results are shown schematically in the accompanying figure.



Azeotropic system.

The vapor curve KLMNP gives the composition of the vapor as a function of the temperature T, and the liquid curve KRMSP gives the composition of the liquid as a function of the temperature. These two curves have a common point M. The state represented by M is that in which the two states, vapor and liquid, have the same composition x_B^a on the mole fraction scale. Because of the special properties associated with systems in this state, the Point M is called an azeotropic point and the system is said to form an azeotrope. In an azeotropic system, one phase may be transformed to the other at constant temperature, pressure and composition without affecting the equilibrium state. This property justifies the name azeotropy, which means a system which boils unchanged.

AZIDES. The salts of hydrazoic acid are termed *azides*. Metallic azides can be prepared from barium azide and the metal sulfate, or from potassium azide and the metal perchlorate.

Soluble azides react with iron(III) salt solutions to produce a red color, similar to that of iron(III) thiocyanate. Sodium azide is not explosive, even on percussion, and nitrogen may be evolved upon heating. With iodine dissolved in cold ether, silver azide forms iodine azide (IN_3), a yellow explosive solid.

Sodium azide is a slow oxidizing agent. It has a selective action in inhibiting the growth of gram-negative organisms. It has been used as a component in selective media such as azide glucose broth or azide blood agar base for the isolation of mastitis and fecal *streptococci.*

A number of alkyl and aryl azides are known, such as CH_3N_3, $C_2H_5N_3$ and $C_6H_5N_3$. The nonmetallic inorganic azides include ClN_3, an explosive gas, BrN_3, an orange liquid, mp −45°C, and IN_3, a yellow solid, decomposing above −10°C. The gas FN_3 is more stable than ClN_3, decomposing only slowly at room temperature.

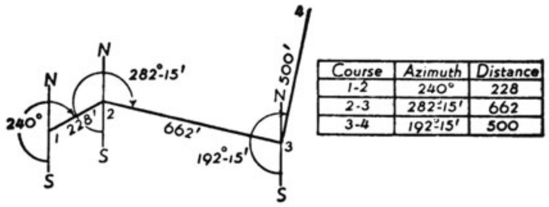
Lead and silver azides are widely used as initiating, or primary explosives because they can be readily detonated by heat, impact, or friction. As such, these materials, particularly lead azide, are used in blasting caps, percussion caps, and delay initiating devices. The function of the azides is similar to that of mercury fulminate or silver fulminate.

AZIMUTH (Astronomy). That coordinate of the horizontal coordinate system of a celestial object which is measured in the plane of the horizon to the point where the vertical circle of the object cuts the horizon, from the south to the right (west) through 360°. Astronomical azimuth may be computed by solving the astronomical triangle, provided three other parts are known. In most cases, the latitude of the observer, the hour angle, and declination of the object are the known parts. In case the longitude of the observer is not accurately known, the altitude of the object may be obtained and combined with latitude and declination for computing azimuth. Several of the terms used here are described elsewhere in this encyclopedia.

AZIMUTH (Navigation). The horizontal direction of a celestial point from a terrestrial point, expressed as the angular distance from a reference direction, usually measured from 0° at the reference direction clockwise through 360°. An azimuth is often designated as true, magnetic, compass, grid, or relative as the reference direction is true, magnetic, compass, grid north, or heading, respectively. Unless otherwise specified, the term is generally understood to apply to true azimuth, which may be further defined as the arc of the horizon, or the angle at the zenith, between the north part of the celestial meridian or principal vertical circle and a vertical circle, measured from 0° at the north part of the principal vertical circle clockwise through 360°. See also **Navigation.**

AZIMUTH (Surveying). The terrestrial azimuth of a mark is usually determined with an altazimuth instrument or surveyor's transit. The difference in azimuth between the vertical circle through some celestial object and the mark is measured and combined with the known azimuth of the object to obtain the azimuth of the mark. The object most commonly used for this purpose is Polaris (the North Star), since this star is within 1° of the pole of rotation, and its azimuth changes very slowly with time. Tables are published in the *Nautical Almanac*, and a variety of other places, which give the azimuth of Polaris in terms of local date and time. When using the North Star for ordinary surveying purposes, the local time is needed only to within about 5 minutes; but for precise geodetic work, the time must be known to within a few seconds.

Surveyors frequently run a traverse using azimuths and distances. The plotting of a traverse by this method is shown in the accompanying figure.



Course	Azimuth	Distance
1-2	240°	228
2-3	282°-15'	662
3-4	192°-15'	500

Plotting of a traverse.

AZINES. The products of the reaction between an aldehyde or a ketone with hydrazine are termed *azines*. A number of dyestuffs and complex members of the pyridine family of compounds also are termed *azines*. See also **Pyridine.**

AZO AND DIAZO COMPOUNDS. Characteristically, these are compounds containing the group —N:N— (azo) or > N⋮N (diazo). They are closely related to the substituted hydrazines. The N_2 group may be covalently attached to other groups at both ends, as in the azo compounds, or at only one end, as in the diazo compounds or diazonium salts. Although organic chemistry furnishes the most numerous examples, many inorganic azo compounds also exist.

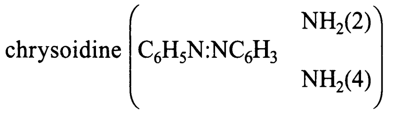
Compounds related to aniline, either directly or by oxidation, and to nitrobenzene by reduction, are numerous and important. When nitrobenzene is reduced in the presence of hydrochloric acid by tin or iron, the product is aniline (colorless liquid); in the presence of water by zinc, the product is phenylhydroxylamine (white solid); in the presence of methyl alcohol by sodium alcoholate or by magnesium plus ammonium chloride solution, the product is azoxybenzene (pale yellow solid); by sodium stannite, or by water plus sodium amalgam, the product is azobenzene (red solid); in the presence of sodium hydroxide solution by zinc, the product is hydrazobenzene (pale yellow solid). The behavior of other nitro-compounds is similar to that of nitrobenzene.

Diazonium salts are usually colorless crystalline solids, soluble in water, moderately soluble in alcohol, and when dry are violently explosive by percussion or upon heating.

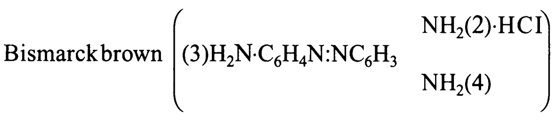
The simplest azo-dyes are yellow, but by increasing the number of auxochrome groups or by increasing the percentage of carbon, the color darkens to red, violet, blue, and in some cases brown. Naphthalene residues darken to red, violet, blue and finally black. These aminoazo-dyes, together with the hydroxyazo-dyes (containing auxochrome hydroxyl-group —OH), are generally only slightly soluble in water. In order that the dye may be soluble it is desirable that it contain one or more sulfonic acid groups —SO_2OH. This group may be introduced either by treating the dye with concentrated sulfuric acid, or by using sulfonic acid derivatives in preparing the dye, e.g., methyl orange, sodium dimethyl-*para*-aminoazobenzene-*para*-sulfonate

$$\mathrm{(4)(CH_3)_2NC_6H_4N{:}NC_6H_4SO_2ONa(4)}$$

from dimethylaniline and diazotized sulfanilic acid (*para*-amino-benzene sulfonic acid, $\mathrm{(1)H_2N{\cdot}C_6H_4{\cdot}SO_2OH(4)}$, and then the sodium salt is made from the product. Other azo-dyes are



$$\text{chrysoidine}\left(\mathrm{C_6H_5N{:}NC_6H_3}\begin{matrix}\mathrm{NH_2(2)}\\ \\ \mathrm{NH_2(4)}\end{matrix}\right)$$



$$\text{Bismarck brown}\left(\mathrm{(3)H_2N{\cdot}C_6H_4N{:}NC_6H_3}\begin{matrix}\mathrm{NH_2(2){\cdot}HCl}\\ \\ \mathrm{NH_2(4)}\end{matrix}\right)$$

AZURITE. This mineral is a basic carbonate of copper, crystallizing in the monoclinic system, with the formula $Cu_3(CO_3)_2(OH)_2$, so called from its beautiful azure-blue color. It is a brittle mineral with; a conchoidal fracture; hardness, 3.5–4; sp gr, 3.773; luster, vitreous, color and streak, blue; transparent to translucent. Azurite, like malachite, is a secondary mineral, but far less common than malachite. It is formed by the action of carbonated waters on compounds of copper or solutions of copper compounds.