E

EAGLE (*Aves, Falconiformes*). Large birds of prey with strong hooked beaks, large curved claws, and powerful flight. The American golden eagle, *Aquila chrysaetus*, is an example of the typical members of the group, which also includes the harpy eagles, hawk eagles, harrier eagles, sea eagles, and others. The bald eagle, *Haliaeetus leucocephalus*, national bird of the United States, is one of the sea eagles.

The golden eagle is a land eagle, brownish in color with purple glossy plumage. The head and neck are brown with filtered yellow. The tail is of several shades of brown and is about 3 feet (0.9 meter) in length. Wingspread is about 7 feet (2.1 meters). The legs are fully feathered. The bird ranges from Canada to Mexico, but is relatively rare in the United States. The golden eagle is found to some extent in England and even more so in Scotland. Nests are on high cliffs. Eggs are from 1 to 3 and are spotted. See accompanying photo.

Golden eagle. (*John H. Gerard from National Audubon Society.*)

The so-called bald eagle is misnamed because of white feathers which make it appear bald. The size is roughly the same as that of the golden eagle. This bird nests on cliffs or in high tree tops. Preferred dietary items are mainly fish which it finds in nearby lakes and rivers. This and related species of sea eagles are found in numerous locations throughout the world.

Bald Eagle Legislation. One of the first wildlife protection actions taken by government was the Act of the U.S. Congress on behalf of the bald eagle several decades ago. Because of earlier wide usage of DDT (dichlorodiphenyltrichloroethane) as an agricultural control chemical, the bald eagle faced almost certain extinction. Although the principal effect of DDT was that of destroying the native habitats of the bird, the substance also caused thinning of the egg shells of the birds, which severely reduced successful hatching and maintenance of the eagle population. Coupled with threats of the carcinogenicity of DDT to human populations, DDT was banned for use in the United States in 1972.

Although, as of the early 1990s, the eagle still is considered an endangered species, the bird has accomplished a significant return in numbers, and some professionals now feel that the bald eagle could be reclassified as a "threatened" rather than an "endangered" species. The legal differences between these two classifications are explained in article on **Endangered Species.**

Legislation has been quite strict. For example, shooting an eagle in the United States may draw a 1-year prison sentence and a $5,000 fine. Anyone finding a dead eagle is legally required to turn it over to the Fish and Wildlife Service. These regulatory actions have produced good results thus far in terms of "saving" the bald eagle, and most likely were aided by the symbolic nature and general public admiration of the eagle as a desirable creature to have in our environment.

In the 1700s, experts estimate that up to 50,000 eagles populated various areas of the United States. In the 1960s, it had fallen drastically to about 800 nesting pairs. By 1991, attributed to protective measures, this number had increased to about 2,600 pairs in the lower 48 states. Prior to the imposition of protective measures, the bird had virtually disappeared from the Great Lakes region and the eastern United States, but now bald eagle concentrations again can be noted in most of these areas. During winter months, thousands of bald eagles took refuge below dams on the Mississippi River, where fish and waterfowl were plentiful, and factors other than agricultural chemicals may be affecting a slower return of the bald eagle to these areas. An assessment of the eagle's predicament prior to passage of the Bald Eagle Act is given by J. W. Grier (North Dakota State Univ.) in *Science*, **218**, 1234 (1982). The current, improved situation is well reported by C. Casey in *American Forests*, 24 (November/December 1990). The biologist M. Amijo (Tahoe National Forest) observes, "Anytime we find a bald eagle nest, the objectives for timber harvest in that area change completely. No longer are we just growing timber—we're growing eagle nesting habitat. We write a silvicultural prescription for that stand that meets the needs of the eagle." Through long-range planning, timber stands now are managed with the target of creating bald eagle habitats that will be useful over the next century or two." As pointed out by biologist P. Detrich (Fish and Wildlife Service), "Bald eagles are relatively easy to manage in the context of forestry. We're not out there harvesting their forage habitat. They depend on lakes and wetlands for food, and they're not directly in the path of the use of forest resources, like spotted owls are. While some other animal species require old growth stands left alone, you can actually improve eagle habitat by doing a little logging."

Of the principal geographic regions of the United States, the "comeback" of the bald eagle has been best in the Pacific, Southwestern, Northern, and Chesapeake Bay locations. The slowest recovery has occurred in the Southeast.

Eagles continue to thrive in Alaska, where the species has escaped listing as endangered. In Alaska, the birds have an excellent supply of salmon and steelhead trout upon which they prey. In the lower states, the greatest concentration of bald eagles is in south-central Oregon and northern California. Annually, a Klamath Basin Bald Eagle Conference Watch takes place and draws large crowds for

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

"eagle watching." It is reported that, during most of January and February, hundreds of the birds can be observed. The "conference" takes place between February 15 and 17. More detail can be obtained from the Department of Fish and Wildlife, 1400 Miller Island Road West, Klamath Falls, OR 97603.

Harrier. This is a moderately large eagle. There are several species, mostly limited to Africa, but one species extends to Asia and Europe. Harrier is also a term used to describe a group of hawks. These hawks are found on all continents. They are slender birds with long wings, and in general are useful as destroyers of reptiles and rodents. The marsh hawk, *Circus hudsonius*, is a North American harrier.

Honey Buzzard. This bird is related to the eagles and is named from its habit of robbing the nests of bees and wasps and eating the larvae. One species lives in Europe and Asia and others in the Oriental region.

Harpy. This is a large, crested eagle related to the true buzzards. Several species range from Mexico to Patagonia. They harpy is reputed to be the most powerful of the eagles. One species in the Philippines (*Pithecophaga jefferyi*) is known as the monkey-eating eagle. The crowned harpy (*Harpy haliaetus*) is found in southern Brazil.

Kite. This is a large bird of prey related to the eagle. The bird is found on all continents. Three species, the swallow-tailed kite (*Elanoides forficatus*), white-tailed kite (*Elanus leucurus*), and the Mississippi kite (*Ictinia mississippiensis*) are found in North America. The black kite (*Milvus nigrans*) and the black-winged kite (*Elanus caeruleus*) are found in Europe.

Secretary Bird. This bird (*Sagittarius serpentarius*) is a remarkable African bird related to the eagles and vultures. The bird is about 4 feet (1.2 meters) tall, with long legs, and is largely terrestrial in habits. It walks and runs very rapidly and is also a strong flier on the relatively rare occasions when it takes to the air.

An excellent article on the recovery of the eagle in North America is given by Peter L. Porteous and Joel Sartore (*Nat'l. Geographic*), 42 (November 1992).

See also **Falconiformes.**

EAR. See **Hearing and the Ear.**

EARTH. Planet *Earth* is one of nine planets that make up the solar system. Each planet is described separately in this *Encyclopedia,* and the solar system is summarized in the article on **Planets and the Solar System.**

The Earth is a slightly oblate sphere (flattened at the poles) that is comprised of several contrasting layers of various materials which, moving outwardly from the center of the planet, are in order: (1) the solid inner and the fluid outer *core*, with a transition zone in between; (2) the lower and upper *mantle*; (3) the *lithosphere*, which supports the *crust* upon which people and other life forms, including plants, insects, and fishes, exist; and (4) a gaseous and vaporous *atmosphere*, also consisting of several identifiable layers, which interface the sphere with outer space. The atmosphere, comprised mainly of nitrogen and oxygen (air), is described in the article on **Atmosphere (Earth).** Approximately three-fourths of the Earth's surface is covered with liquid water (oceans) and solid water (ice caps). See also **Ocean.**

The Earth is the only planet in the solar system or, in fact, in all of the cosmos known to support life as we currently define it. Thus, while there are numerous similarities among the planets, in terms of life, the Earth is unique.

Only during the past few decades have people been able to observe their planet as an entity—that is, to see Earth in one full and continuous panorama as imaged from an artificial satellite. See Figs. 1 through 3. Earlier, partial views had to be made from high-flying aircraft and presented in the form of mosaics. See **Satellites (Scientific and Reconnaissance).** Today, the Earth can be viewed by all manner of imaging and sensing instruments, thus dramatically expanding scientific knowledge of the planet. Weather satellites, in a very practical way, now permit the observation of cloud coverage and patterns across entire continents and oceans. Satellite-based thermal, radiation, and other sensors have markedly improved the reliability of weather forecasting. See **Weather Technology.**

Fig. 1. The Earth as seen from outerspace. Astronauts described the Earth as "a colorful island in the vast sea of blackness." The Northern Hemisphere: (1) United States West Coast; (2) Pacific Ocean; (3) Atlantic Ocean; (4) Gulf of Mexico.

Fig. 2. The Southern Hemisphere as seen from outerspace. The South Polar area is visible under partial cloud cover.

Earth is the third closest planet to the sun, with Venus being closer and Mercury being the closest. Then, further outward from the sun and beyond the Earth are the other six planets—in order, Mars, Jupiter, Saturn, Uranus, Neptune, and Pluto. Figure 4 is included here to dramatize just how great the planetary distances are. These distances are further realized when considering how long it takes for a space probe launched from the Earth to reach the region of a given planet. For example, *Voyager 2*, launched from Earth on September 1, 1977, required 22 months to reach Jupiter, a total of 49 months to reach Saturn, a total of 137

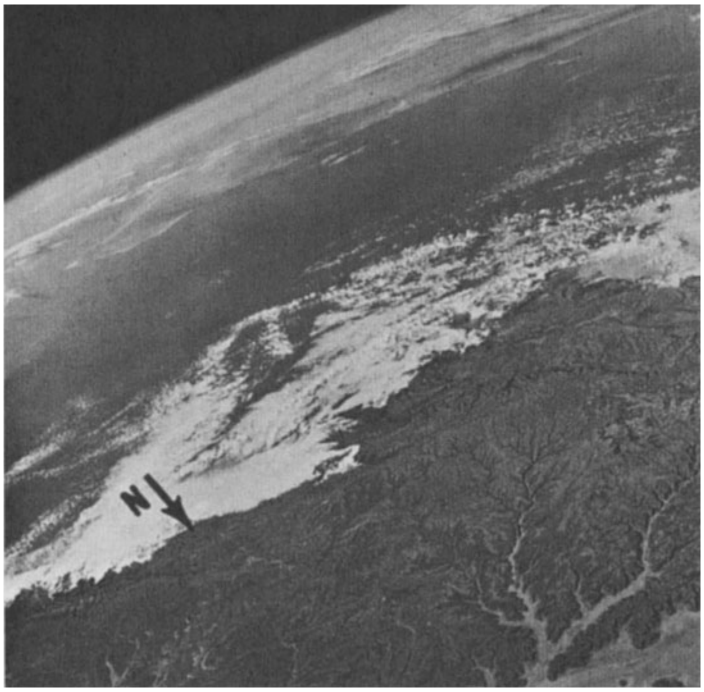


Fig. 3. Small portion of the Earth as viewed from an altitude of 100 miles (161 kilometers). View shows the Hadramaut plateau, the Gulf of Aden, and Somalia.

months to reach Uranus, and a total span of 145 months (12 years, 1 month) to fly by Neptune. See also **Voyager Missions.**

In terms of size, the Earth is larger than four of the planets (Pluto, Mercury, Mars, and Venus), but is much, much smaller than Juipter, Saturn, Uranus, and Neptune. See Fig. 5.

The Earth has a single satellite (the *Moon*). Mercury and Venus have no satellites, but all other planets have from one to two or more satellites, with Jupiter leading with over a dozen moons.

The solar system, including the Earth, is believed to be part of a spiral galaxy M31 (the Andromeda Galaxy), which, with a few dozen other small galaxies, constitute what is known as the *Local Group* and part of the Milky Way.

For a number of years, some cosmologists have voiced an assumption, based upon statistical probability, that other solar systems, supported by other stars, may have planetary bodies, some of which may be similar to the Earth. None of these postulated systems and bodies have been "discovered," let alone confirmed. But, on the basis of such hypothecation, a search has been underway to "communicate to and from" speculative other intelligences located elsewhere in the cosmos.

Prior to describing some of what is known about the Earth's interior and other earthly features, Tables 1 through 9 summarize useful facts pertaining to geodetic parameters, geometric parameters, major physical features (continents, oceans, lakes, and rivers), meteorological extremes (temperature and precipitation), and major floods, tidal waves, earthquakes, volcanoes, and eruptions. Also there are separate articles on **Earth Tectonics and Earthquakes,** and **Volcano.**

The age of the Earth has been variously estimated, with a figure of something less than 4.6 billion years sometimes used. Such figures are based upon theoretical calculations and extrapolations made by some cosmologists, but on which solid agreement does not apply. See **Cosmology.** In any event, when plotted against any humanly comprehensible time scale, the Earth is very, very old. The word *ancient* simply does not adequately convey a time scale of this magnitude.

Later, in this article, the topic of the presently changing state of the Earth, commonly referred to in contemporary literature as "Global Change," is addressed. As will be developed in this area, it is becoming increasingly difficult to sort opinions and facts.

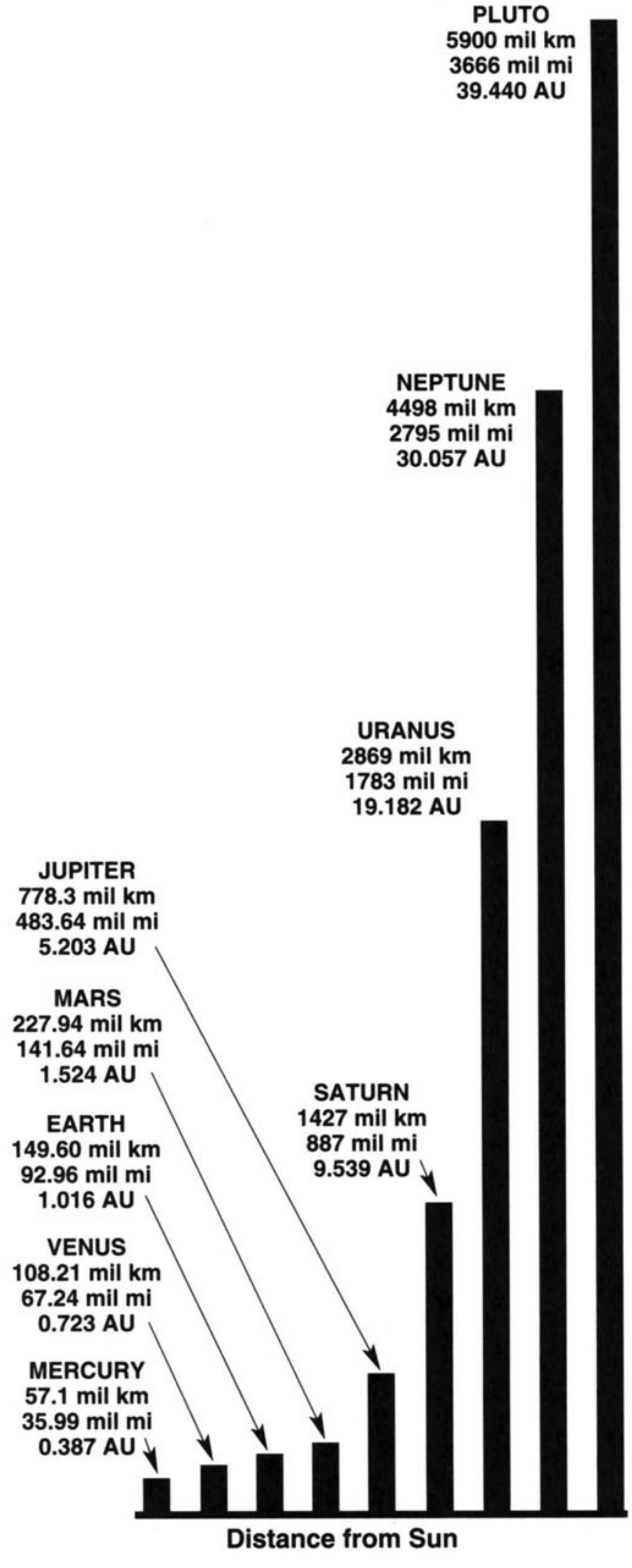


Fig. 4. Comparison of the distances of the nine planets of the solar system from the sun. Distances are given in millions of kilometers and miles and in Astronomical Units (AU).

Geophysics

Geophysics is the physics of the Earth and the space immediately surrounding it and the interactions between the Earth and extraterrestrial forces and phenomena. Geophysics consists of a number of interlocking sciences dealing with physical properties of the Earth, its inte-

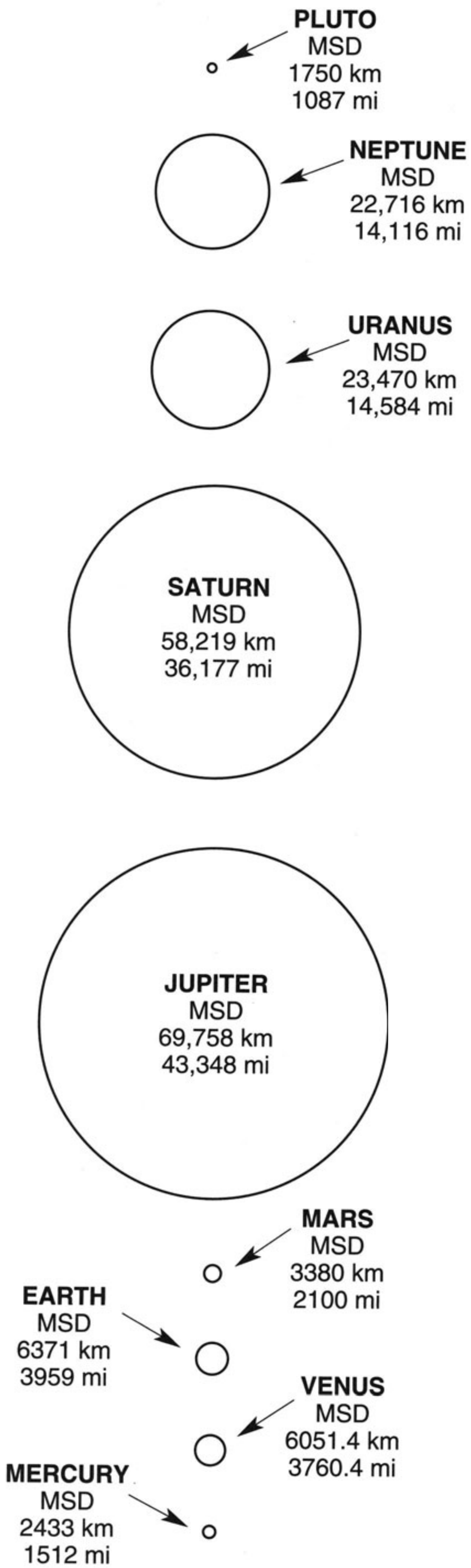


Fig. 5. Comparison of dimensions of the nine planets of the solar system. Figures given in kilometers and miles are for the mean semidiameter (MSD) of each planet shown.

rior and atmosphere, its age, motions and paroxysms, and their practical applications. All of these sciences use the methods of physics for measurements and analysis. From observational material, often of an indirect nature, attempts are made to derive abstract models of states and processes through advanced mathematical concepts and, in some cases, through statistical relations.

In essence, geophysics is the study of the Earth as a planet—with three basic divisions of activity—the solid Earth, the atmosphere and hydrosphere, and the magnetosphere (solar-terrestrial physics). As a convenience, the term geophysics sometimes has been extended to similar studies of the planets and their satellites. However, for extraterrestrial studies, there are more specific terms, such as selonography which relates to the moon.

Geophysics may be described as an ancient science. In its early stages, it was developed by the Greeks who attempted to determine the shape and size of the Earth (Eratosthenes, 275-194 B.C.). Among its most illustrious contributors have been Galileo Galilei (1564–1642); Sir Isaac Newton (1642–1727), who dealt with the motions of the Earth and its gravitational field; Karl Friedrich Gauss (1777–1855), who developed the theory of the magnetic field; and Vilhelm Bjerknes (1862–1951), who laid the foundation for the hydrodynamic theories of the atmosphere and the oceans. The roster of distinguished scientists who contributed to this field during the 20th century includes: L. Vegard (polar aurora); Sidney Chapman (aeronomy); C. G. Rossby (meteorology); H. U. Sverdrup (oceanography); Sir Harold Jeffreys; F. A. Vening-Meinesz (structure of the Earth); and B. Gutenberg and J. B. Macelwane (seismology).

A major series of milestones toward the advancement of scientific knowledge of the Earth commenced, on the basis of international cooperation, in the late 19th century, with the First International Polar Year (1882–1883), followed by the Second International Polar Year (1932–1933), fifty years later. For the International Geophysical Year (1957–1959), over 8,000 scientists of 66 nations collaborated and spawned the more recent ventures, including satellite investigations of the Earth (Skylab and its predecessors), increased exploration of the Antarctic continent, the International Years of the Quiet Sun (IQSY, 1964–1965), and numerous different and subsequent activities. The International Decade of Ocean Exploration (1970–1979) was a very rewarding program. Other international programs proposed or underway, as of the late 1980s, include the International Geosphere-Biosphere Program (IGBP), targeting the global climate, the biosphere, and the biogeochemical cycles of all major nutrients; the World Climate Program, and the International Biological Program. A number of geoscientists have expressed, however, that limiting such programs to a single year or even a decade has the disadvantage of breaking up a complex infrastructure of specialists and procedures involved in projects that naturally are of a much longer-term nature.

Structure of the Earth

One contemporary concept of the structure of the Earth is diagrammed in Fig. 6. An estimate of the various parameters which apply to the numerous layers and regions of the planet are given in Table 1.

Crust (Lithosphere). Discounting the atmosphere and hydrosphere, the crust is sometimes defined as that part of the Earth that is situated above the Mohorovičić discontinuity, defined later. It is on the crust where living creatures exist—a layer of the Earth that accounts for less than 0.2% of the whole planet. The lithosphere consists of three shells: (1) a *stratified sedimentary shell*, composed mainly of sedimentary rocks; (2) a *granitic shell*, distributed only beneath the continents and thinning out at the ocean boundaries; and (3) a *basaltic shell*, the structure of which differs somewhat when situated under continents or under oceans.

Sedimentary Shell. The total volume of the sedimentary shell is about 1.05 billion cubic kilometers, taking into consideration the consolidation of recent sediments, and 900,000,000 cubic kilometers without volcanic rocks, i.e., about 10% of the volume of the crust and 0.1% the volume of the whole Earth. The average thickness of the sedimentary shell is 2.0 kilometers; if the area of the shields not covered by sediment is excluded, the average is 2.2 kilometers.

On the continents, about 75% of the volume of all sedimentary rocks is found in geosynclinal areas and only 25% in the platforms, their average thickness being 10 kilometers and 1.8 kilometers, respectively.

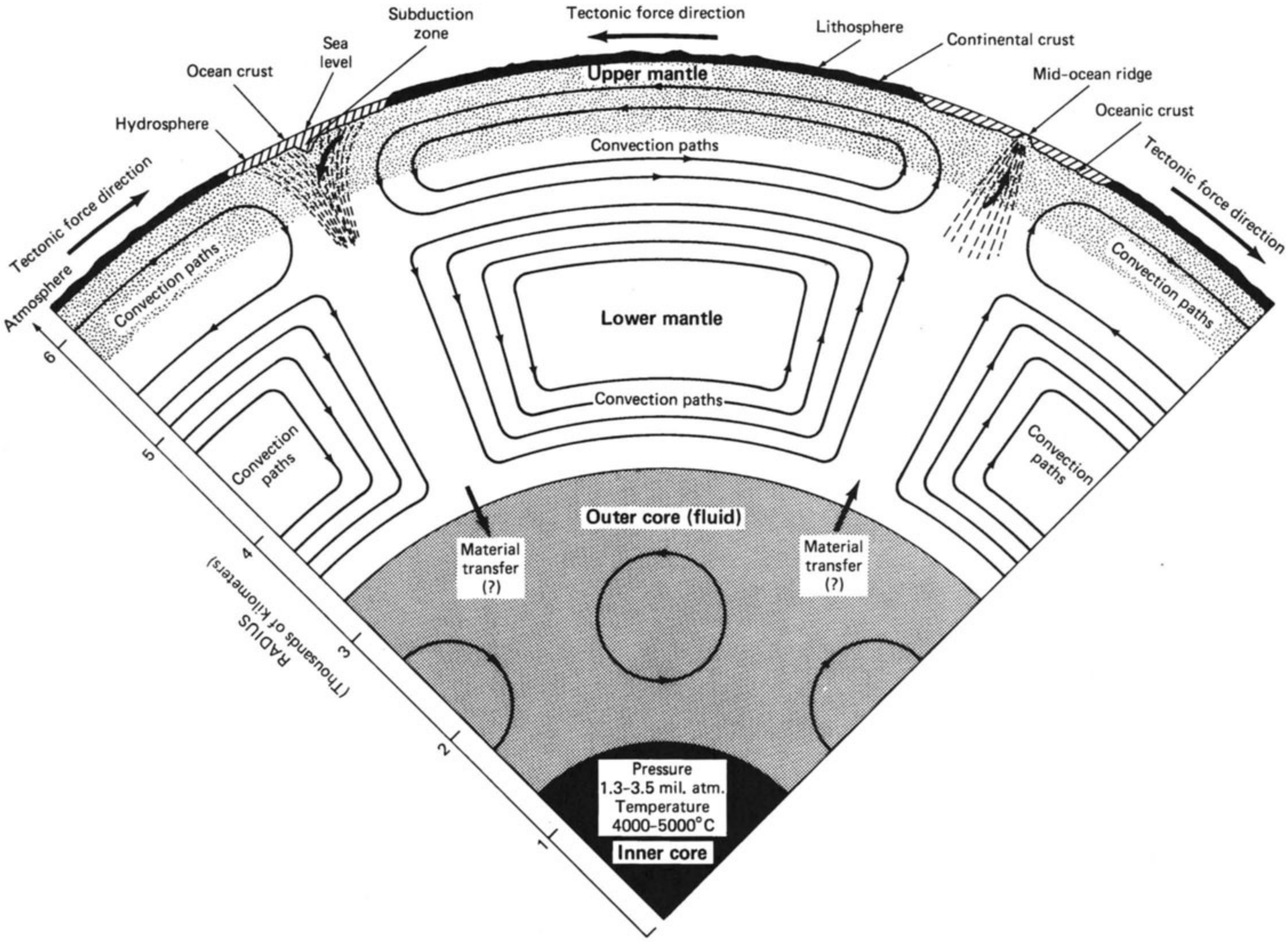


Fig. 6. Tentative, contemporary concept of structure of the Earth. (*After Siever.*)

Clay and shale are the most widespread sedimentary rocks on the continents (42%). Arenaceous, volcanic, and carbonate rocks are approximately equally abundant (20, 19, and 18%). All other rock types, mainly evaporites, comprise about 1%.

Granitic Shell. This shell is restricted to the continents and its volume and mass are approximately 3.6 billion cubic kilometers and 9.8 $\times 10^{24}$ grams, respectively. Acidic granitoids and metamorphic rocks are the main rock types of the granitic shell; the basic and ultrabasic rocks make up less than 15% of the shells' volume. These volumetric ratios of rocks lead to the acidic chemical composition of the shell, its typical high content of silica, the concentrations of alkalies K > Na, and of the rare elements (uranium, thorium; rare earths; zirconium, niobium, etc.).

Basaltic Shell. This shell consists of two parts, the continental and the oceanic, differing in structure and apparently in composition. According to one hypothesis, the basaltic shell of the continental crust is formed of strongly metamorphosed rocks of both basic and acid composition, together with a significant portion of magnetic rocks.

The *continental crust* has considerable thickness and a diversity in composition. The more homogeneous oceanic crust is 86% original oceanic theoleiitic basalts and their metamorphic equivalents. These basalts are characterized by a low content of potassium, rubidium, strontium, barium, phosphorus, uranium, thorium, and zirconium, and high ratios of K/Rb and Na/K, which strongly distinguish them from analogous continental rocks.

The *oceanic crust* is essentially characterized by the occurrence of ultrabasic rocks, seen in the zones of deep faulting (mid-ocean rift valleys), these rocks being considered outcrops of mantle material.

In summary, about 64% of the whole crustal volume is continental, or 79% when the shelf and subcontinental (quasicratonic) crust are included. The other 21% is oceanic crust. The average thicknesses of the various crustal types decrease from 43.6 kilometers for continental, to 23.7 kilometers for subcontinental, to 7.3 kilometers for oceanic crust. The average thickness of the entire crust amounts to about 20 kilometers.

The chemical composition of the crust as a whole approaches that of intermediate rocks, though it is impossible to find its close analog among them. In rough approximation, the crust's composition can be described as a mixture of the two prevailing types of rocks: (1) granite, and (2) basalt (geosynclinal basalt plus oceanic theoleiite) at a ratio of 2:3. The average chemical composition of the crust changes with depth from the sedimentary shell to the basaltic shell, with a continuous increase in the content of iron, magnesium, and alumina; and a decrease in the amount of combined water. The contents of alkalis (potassium) and silica first increase from the sedimentary shell toward the granitic; and then decrease toward the basaltic shell of the continents and oceans.

Mohorovičić Discontinuity. Often simply termed the *Moho discontinuity*, this is the boundary surface or sharp seismic-velocity discontinuity that separates the Earth's crust from the subjacent mantle. The discontinuity marks the level in the Earth at which *p*-wave velocities change abruptly from 6.7–7.2 kilometers per second (in the lower crust) to 7.6–8.6 kilometers per second, or an average of 8.1 kilometers per second (at the top of the upper mantle). The depth of the Moho discontinuity varies from about 5–10 kilometers beneath the ocean floor to about 35 kilometers below the continents. Its depth below

TABLE 1. APPROXIMATION OF EARTH'S INTERNAL LAYERING AND VALUES OF SOME PHYSICAL PARAMETERS

Depth (km)		Fraction of Volume[a]	Density (g/cm^3)	Pressure (10^{12} dyn/cm^2)	Gravity (cm/sec^2)	Rigidity (10^{12} dyn/cm^2)	Characteristics of P, S Velocities[b]	Features
0	Crust	0.0155					Complex	Heterogeneous
33	Moho discontinuity							
			3.3	0.01	985	0.6		
	Upper mantle (Region B)	0.1667					Normal gradient	Probably homogeneous
410								
	Upper mantle (Region C)	0.2131					Greater than normal gradients	Transition layer
1000			4.7	0.4	995	1.9		
	Lower mantle (Region D′)						Normal gradients	Probably homogeneous
2700		0.4428						
	Lower mantle (Region D″)						Gradient near zero	Transition layer
			5.7	1.3	1030	3.0		
2900	W-G Discontinuity							
			9.7	1.3	1030	0.0		
	Outer core	0.1516					Normal P gradient	Homogeneous fluid
4980								
	Transition region	0.0028	(12.5)	3.2	(500)	(0.2)	Negative P gradient	Transition layer
5120								
	Inner core	0.0076					Smaller than normal P gradient	Solid
6370[c]			(13.0)	3.7	0	(1.3)		

[a]Volume ratio, crust:mantle:core = 1:51:10.
[b]P = pressure wave; S = shear wave.
[c]The value of 6,370 km depth refers to center of earth.

some mountain regions may reach 70 kilometers. It is reasoned that the discontinuity represents a chemical change between the basaltic materials above to periodotitic or dunitic materials below, rather than a phase change (basalt to eclogite). The discontinuity should be defined in terms of seismic velocities alone until more fundamental findings are made. The discontinuity is variously estimated to be between 0.2 and 3 kilometers thick. The discontinuity is named after Andrija Mohorovicić (1857–1936), the Croatian seismologist who discovered the phenomenon.

The Mantles. For many years, that immense layer or region of the Earth's interior that lies between the Moho discontinuity and the core of the planet was referred to simply, in the singular, as the mantle. Over the years, possible layers among the mantle were proposed and discussed, but it has only been comparatively recently that there is broad acceptance of the existence of two layers in the mantle, an *upper mantle* (sometimes called *asthenosphere*), which intersects with the Moho discontinuity, on its upper side; and a *lower mantle*, which intersects with the upper mantle and on its bottom side, with the outer core approaching the center of the Earth. Much has been learned, and even more speculated, concerning the materials and characteristics of the upper mantle. There is considerable speculation concerning the lower mantle, how it may interact with the upper mantle and with the outer core and whether or not it may mimic the general behavior of the upper mantle.

When considered together, the upper and lower mantle, these layers account by far for the major portion of the inner volume of the Earth.

As early as 1910, Alfred Wegener proposed that the continents drift. Many years later, the plate tectonics concept was developed and widely accepted, not only for explaining the apparent very slow movement of land masses borne by underlying plates that float on the upper mantle, but also for elucidating the cause of most earthquakes and volcanic activity. The concepts of currents flowing within the upper mantle as plate driving forces had been proposed and studied from time to time, but remained for the late 1970s and early 1980s for more concerted research to find more convincing evidence for the existence of such currents. The logic and scenario followed by several researchers during this latter time frame is most interestingly developed by D. P. McKenzie in a review report (see reference). Unfortunately there is insufficient space available to present many of the details.

D. P. McKenzie and N. O. Weiss (University of Cambridge), both authorities on tectonics, found that the mass of data collected pertaining to plates (for example, where they are located, approximate rates of movement, reactions resulting from such movements, their size, on the average 100 km thick, and other factors) simply was insufficient to help in providing leads pertaining to probable convection currents operating in the underlying mantle. It was, of course, almost self-evident that some underlying force had to be present and in sufficient magnitude to cause the plates to move. At the estimated temperatures and pressures existing in the upper mantle, as learned from observing volcanoes, earthquakes, and other tectonic related events, it seemed logical to assume that the mantle was fluid and that some force, most likely flow resulting from temperature and/or pressure gradients, was present. It is interesting to note that a number of geophysicists for several years had insisted that the mantle cannot flow like a liquid, at least when taking into consideration the probable composition of the materials that make up the mantle. A number of researchers, using various techniques, such as laboratory and computer modeling of fluid dynamic processes, the use of satellite gravity field and surface deformation measurements, precise seismic investigations, and isotope ratio studies, have contributed to the establishment of the mantle convection currents concept.

For example, McKenzie and Weiss found from laboratory fluid dynamics studies, where flows were measured and observed in experimental tanks and supported by some numerical modeling, that mantle convection probably consists of at least two scales of motion. Transposed to terms relating to the mantle situation, the researchers suggested that a small-scale circulation with a distance of about 1500 km between cold, sinking regions be superimposed on a larger-scale circulation that returns materials from a trench to a ridge. Both of these processes are now well understood by investigators in plate tectonics. McKenzie and Weiss believed that the two-scale model of cirulation in the mantle could reconcile the geophysical observations with the behavior of convecting fluids as observed in the laboratory. A problem remained, however—the existence of the small-scale circulation could not be directly observed. This remained pending the availability of satellite instrumentation that could be used to map the Earth's gravity field very accurately. As observed by McKenzie, numerical calculations had shown that upwelling regions of convection should be associated with small, positive gravitational anomalies and that flow should push up the Earth's surface. Plates are too thin to have much effect on either the gravity field or the surface deformation. Therefore, it was reasoned that if the gravity field and the surface deformation can be accurately mapped, it should be possible to "see through" the plates and map the convective circulation under them.

McKenzie, in cooperation with B. Parsons (Massachusetts Institute of Technology), A. Watts (Lamont-Doherty Geological Observatory of Columbia University), M. Roufosse (Center for Astrophysics of the Harvard College Observatory and the Smithsonian Astrophysical Observatory) were indeed able to map the small-scale convection cells in the mantle. Even though the maps confirm the general features of the two-scale model previously mentioned, at least two further questions remained unresolved.

As previously mentioned, some geophysicists did not feel that the mantle can flow like a liquid mainly based upon materials considerations. It became evident to some investigators, however, that the answer may be found in what metallurgists term *creep*. (This is evidence of the multidisciplinary nature of geophysical research.) The flow behavior of materials close to their melting point had been examined in connection with jet engines and nuclear reactors. Investigators learned that under these conditions all crystalline materials flow under any stress, regardless of how small it may be. It was found that high-temperature creep differs in many respects from low-temperature creep. Incidentally, the concept of solid rock at high temperatures has resolved a number of puzzling situations in geology.

Gravitational and bathymetric data contributed to confirming circulation of the mantle. See Fig. 7. Although the investigations of residual depths and gravitational anomalies made it possible to map rising and sinking regions, they provided little information about the depth to which the circulation extends and no information concerning its evolution with time. The researchers then turned to data on isotope ratios as possible clues. In the early stages of the formation of the Earth, certain elements, such as strontium and rubidium, were concentrated in the Earth's crust, but were depleted in the upper mantle because the Sr and Rb ions, for example, did not easily fit into the lattice framework of most minerals present in the upper mantle. However, relatively more Rb concentrated in the crust than Sr. Similarly, relatively more neodymium (Nd) concentrated in the crust than samarium (Sm). Applying the knowledge of how these various radioactive isotopes decay, one can ultimately determine the migration of materials, and hence deduce the transfer of materials that had to occur to effect such results. See Fig. 8. The research and logic followed by the investigators is further detailed in the paper by McKenzie, who, in 1983, summarized as follows: "Twenty years ago many earth scientists considering the evidence for or against continental drift were consciously or unconsciously thinking in terms of a static model of the earth. This situation changed completely with the general acceptance of sea-floor spreading and plate tectonics. The effect on the study of the dynamics of the mantle was particularly profound, since plate tectonics established the existence of mantle convection without providing much information about the forces involved. Some of the first attempts to understand mantle dynamics limited the circulation to the plate motions and a return flow that carried the mantle material from trenches to ridges. The dynamic models and observations of the gravity field have now clearly shown that much of the convective circulation is not related to the movements or boundaries of the plates. I believe we now understand the outlines of the dynamics of the upper mantle; the challenge is to discover how the more massive lower mantle behaves."

Surface Clues of the Mantle. The mantle beneath the continents is usually buried beneath many kilometers of continental crust. In the Ivrea zone of northern Italy, geological and geophysical data indicate

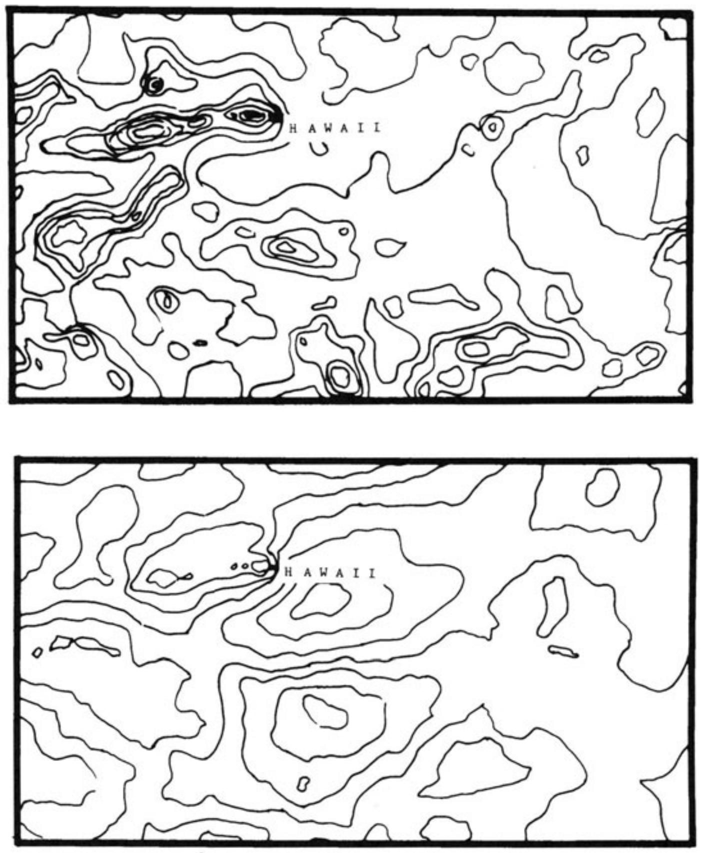


Fig. 7. Black-and-white facsimiles of color enhanced bathymetric anomaly (top) and gravitational anomaly (bottom) maps of portion of Pacific Ocean. These data were important in developing concepts pertaining to mechanisms involved in tectonics and, notably, a better understanding of the upper mantle of the Earth, including the presence of convection currents. Both maps have been smoothed to eliminate fluctuations less than 500 km. The bathymetric anomaly is plotted at "residual depth," which is the difference between the depth that can be attributed to the contraction of the oceanic plate as it cools and the observed depth. The gravitational anomaly is observed as a fluctuation in the height of the sea surface, measured by radar altimeters carried aboard satellites. To investigators, the maps revealed that where the sea surface tends to bulge, the residual depth is positive and that both features are expected above an upwelling region in the mantle. Similarly, the sea surface tends to be depressed where the residual depth is negative, as expected above a downwelling region. The map is projected so that the motion of the plate with respect to the mantle is always to the left over the entire region. This motion generates a small but detectable elongation of the anomalies in the direction of the motion, causing them to appear like ellipses whose long axes run horizontally across the diagrams. (*After McKenzie.*)

that the rocks have been thrust upward and exposed to view. Ultrabasic rocks outcrop in the mountains near the town of Finero.

The Red River near Yuanjiang in Yunnan, China, flows along a marked crustal discontinuity between underformed late Precambrian rocks to the north and deformed and metamorphosed Mesozoic rocks to the south. Geological discontinuities of this sort can be used to identify ancient plate boundaries.

At Dead Horse Point, Utah, large-scale vertical movements are clearly demonstrated in the western United States, where major uplift has produced some of the most spectacular scenery. Other movements, often oscillatory in character with rates that sometimes are on the scale of centimeters per year, are less obvious. The relationship between these vertical movements and the simple plate tectonics model is not clear.

These rare locations are illustrated in article by Drake and Maxwell (see reference).

Deep Holes. Ironically, it is many orders of magnitude easier to probe the depths of outer space than the interior of the Earth and, consequently, a large percentage of our knowledge pertaining to the inner Earth is inferential. Deep mines and deep exploratory drill holes have provided a glimpse of the top 10 km or so of the continental crust.

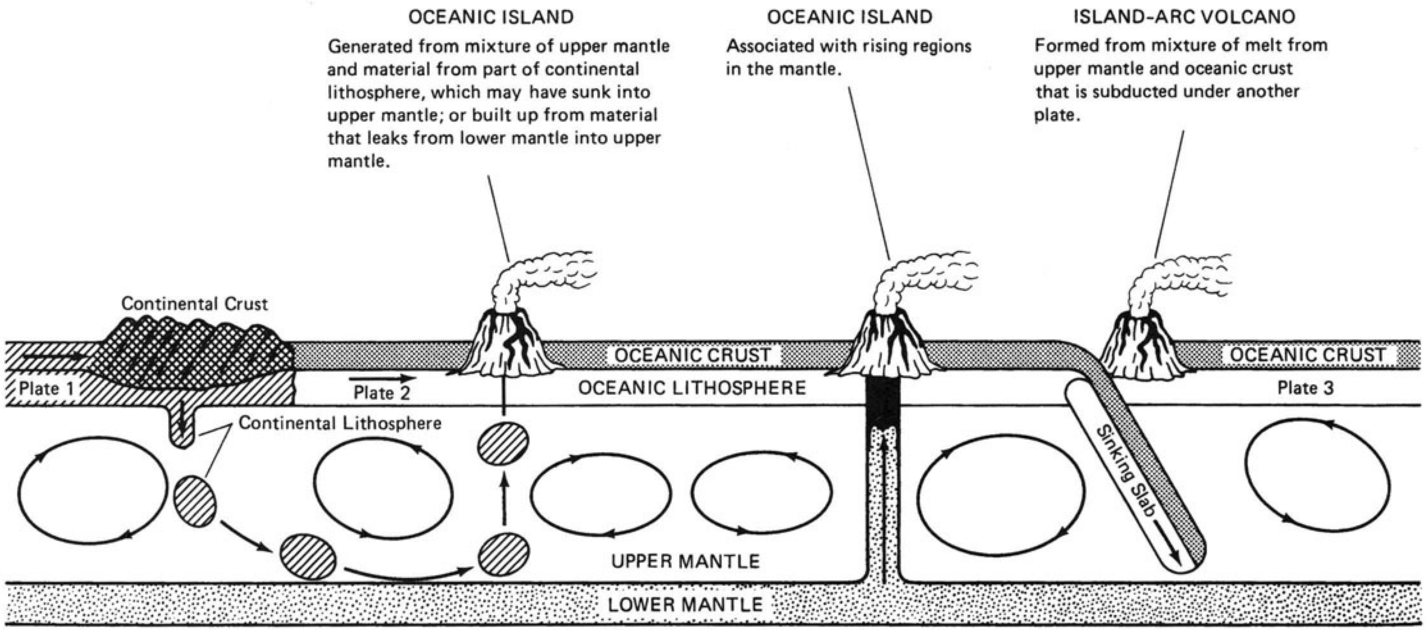


Fig. 8. Highly schematic diagram depicting probable scenarios for formation of continents, oceanic islands, and island-arc volcanoes. Relationship between upper mantle and oceanic and continental lithosphere, continental crust, oceanic crust, and patterns of circulation as proposed for the upper mantle are also shown. As reported by McKenzie, the origin of material erupted onto volcanic islands can be inferred from the isotopic ratios of certain elements. In the early phases of the Earth's history, some specific elements, such as strontium (Sr) and rubidium (Rb) were concentrated in the Earth's crust and thence depleted from the upper mantle, these actions deduced from the fact that the ions of these elements do not easily mesh into the lattice framework of common minerals. Data indicate that the concentration of Rb in the crust was relatively greater than that of Sr. Because all isotopes are chemically equivalent, the ratio of Sr 87 to Sr 86 remained the same after differentiation of the crust and upper mantle. However, thereafter, the ratio of Rb to Sr varies because half of the Rb 87 radioactively decays to Sr 87. Thus, radioactivity becomes a clue in tracing how materials are transferred in the formation of the major features of the Earth and ultimately assists in determining the probable convection circulation occurring in the upper mantle. Because of an initial surplus of Rb 87 in the crust, the ratio of Sr 87 to Sr 86 increases most rapidly in the crust, less rapidly in the Earth as a whole, and least rapidly of all in the upper mantle. Thus, the ratio of Sr 87 to Sr 86 becomes the key to this kind of analysis. Two other elements, samarium (Sm) and neodymium (Nd), also were both concentrated in the crust, with relatively more Nd than Sm transferred to the crust. Through radioactive decay, half of the Sm 147 decays to Nd 143. Thus, as in the case of the Sr isotopes, the ratio of the Nd isotopes serves as a similar clue for this type of analysis. Although most of the major features of tectonic activity near the surface of the Earth have been known for several years, tracing materials origins and transfers and hence upper mantle circulation has been enhanced through these isotopic studies and comparisons. (*Sketch is after McKenzie.*)

The *Glomar Explorer* and *Glomar Challenger* have drilled into the oceanic crust. See **Ocean Research Vessels.** Much of the data potentially obtainable from commercially oriented drill holes either is not observed at all or it is not readily available to scientists. In the early 1980s, the National Academy of Sciences established a Continental Scientific Drilling Committee with the objective of advising scientists of opportunities for obtaining samples and for performing various kinds of geophysical observations, such as regional stress determinations.

In addition to the Glomar vessels previously mentioned, a number of scientific-sponsored drilling projects have been undertaken or are currently underway. Others remain in the proposal stage. A large and ambitious drilling program has been in progress in Russia, and just a few years ago, foreign scientists were permitted to visit the Kola hole, located near the Arctic Circle. At that time the hole was 12 km in depth. The drilling derrick is 64 meters (210 feet) high, and an entire scientific and supporting community has been built around it. The Kola hole is only one of several Russian drilling programs, which also include a second superdeep hole now underway. This hole is located at Saatly, near the Caspian Sea and already has reached a depth of well over 8 kilometers. The plans also include a web of geophysical profiles that will connect eleven deep and superdeep boreholes as part of an effort to elucidate the geology of Russia and to identify new mineral resources. At Kola, the drill bit can progress from 2 to 3 meters per hour of drilling. Drill rig automation permits full retrieval and reinsertion of 12 km of drill pipe in 18 hours. This is considered exceptional on ordinary drilling standards.

Drilling projects do not have to be spectacular to gain useful scientific information. The Inyo Domes Project on the eastern edge of the Sierra Nevada just east of Yosemite National Park is an example. Here, scientists were investigating the flat-topped domes which were formed where thick lava had oozed to the surface—in an effort to study how fluid rock makes its way through the brittle upper crust and either quietly flows out or is explosively shattered into ash and strewn across the adjoining land. This project involved slant-drilling to a depth of about 650 m below the center of Obsidian Dome, at which depth the drill bit encountered the dike demonstrating that at least above that depth, the dike rises vertically to the surface. Much geophysical information was obtained from the project. Further findings are reported in the Kerr (1985) reference.

For a number of years, an ambitious continental drilling project in the United States, sometimes referred to as the Southern Appalachian Superdeep Hole, has been under consideration. The 10-kilometer hole would be nearly twice as deep as any drilled through hard, crystalline rock other than the Kola hole. Drilling specialists do not envision many difficult problems. The drilling rig would have to lower and raise a 15-kilometer long, 450,000-kilogram pipe and its attached drilling bit. By comparison, oil and gas strings are only about one-third that weight. The drilling would have to be straighter than that achieved in normal drilling practice—with deviations of less than a few degrees per 100 meters. Straightness is important because abrasion at the slightest of bends would wear out the drill string. Conditions in the lower portions of the borehole are envisioned as comparatively cool (165°C) as compared with about 250°C for a deep gas well. It is also envisioned that all but the uppermost section of the borehole would be sufficiently strong to forego the lining usually required in deep wells.

Deep drilling experience in hard rock in the United States has been quite limited. As reported by Kerr (1984), two boreholes of about 4 km were sunk at Fenton Hill near Valles Caldera in New Mexico as part of the hot dry rock geothermal project of the Los Alamos National Laboratory. The deepest hard-rock hole in North America was an accident, drilled by Phillips Petroleum Company in a search for oil and gas in Arizona. As reported by Kerr, deep drilling programs are also underway in Belgium, France, and Germany.

Some years ago, the Mohole Project had the single target of a deep ocean hole through the Moho discontinuity, an uncompleted task now envisioned as much more difficult than drilling the southern Appalachian hole.

The Cores. From the total mass of the Earth and its moment of inertia in conjunction with seismic information, it has been deduced that the core contains about 1.95×10^{27} grams of material, constituting about one-third of the mass of the earth, while occupying only one-sixth of its volume. Seismological data indicate that the core extends from a depth of about 2900 km (1800 mi) to the center of the earth, which is at a depth of 6370 km (3960 mi). Such data also indicate that the inner core is solid and has a radius of about 1200 km (1920 mi). The inner core is surrounded by a liquid outer core. Pressures in the core are estimated to range between 1.3 and 3.5 million atmospheres and the temperature is estimated to range from 4000 to 5000°C (7200–9000°F).

An interest in what occurs under the surface of the Earth and particularly of what the Earth is like at its center has fascinated people for centuries. Much of the information from the past has been derived from studies of the magnetization of rocks. Inspection of some of the oldest known rocks suggests that the mechanism which generates the geomagnetic field has been in place for at least 3.5 billion years. For many years, the generator of the geomagnetic field has been likened to a dynamo, but only quite recently has further new evidence been obtained to strengthen that concept.

As more is learned of *what* appears to be the main function of the Earth's core, clues are produced about *how* it may function. Thus, much research in recent years has been directed to determine how the geomagnetic field interacts with other phenomena and, in particular, how it interacts with the solar wind to learn about the nature of the Earth's magnetotail. A number of other questions remain. When did the geomagnetic field first appear? How was it started? How has it evolved? Has it changed over the past eons? It is undergoing change, even if change is measured in millions upon millions of years? The why of the Earth's dynamo is elusive.

Traditional View of Geomagnetic Field. Many geoscientists over the years have compared the geomagnetic field with a bar magnet at the center of the Earth, with lines of force looping from the South Magnetic Pole to the North Magnetic Pole. This is a rather limiting comparison, however, because in reality the magnetic field behaves like that of a dipole only near the surface of the Earth. It is now known that the geomagnetic field is distorted by the solar wind. Even before the concept of the solar wind was developed, a connection between disturbances in the geomagnetic field and solar activity (sun spots) had been well established. Space exploratory missions have taught much about the magnetic fields and the effects of solar radiation on the magnetic fields of other planets. Studies of aurora phenomena have revealed a direct connection between the solar wind and the geomagnetic field. This is described in greater detail later in this article.

It is of interest to review briefly the research and lines of reasoning brought to bear on the question of the Earth's core dynamo. For a more penetrating analysis, R. Jeanloz' paper (see reference) is suggested.

The concept that the geomagnetic field generator takes the form of a magnetohydrodynamic machine was proposed by W. M. Elasser (Johns Hopkins University) and E. C. Bullard (University of Cambridge), among others. Jeanloz points out that these processes entail convection in an electrically conducting fluid, with the result that the core acts as a dynamo, maintaining and regenerating the magnetic field. An article on **Magnetohydrodynamic Generator** in this encyclopedia describes the consideration of this technology as an emergency source of electrical energy by electric utilities.

As field lines directed toward the center of the Earth (*poloidal lines*) enter the outer core, they are pulled in the direction of the Earth's rotation. Thus rotation of the solid inner core probably tends to wrap the field lines around the Earth's axis, producing a toroidal component. It is also speculated that the field lines may become-contorted by smaller-scale cyclonic motions that result from the assumption that the core is rotating essentially in synchronism with the rest of the Earth. These cyclonic motions may be likened to the hurricane patterns that arise in the atmosphere. The exact origin and detailed pattern of the contortions remain unknown.

It is interesting to note that in the absence of the dynamo process, the Earth's magnetic field most likely would have died out within 10,000± years and yet it is still performing a few billion years after the Earth was formed.

Matching Core Characteristics with the Geomagnetic Field. Traditionally, the inner core of the Earth has been regarded as consisting mainly of iron (Fe). This assumption agrees well with most seismic data. There are at least two other clues which add to the credibility of Fe in the core: (1) From our present knowledge of electricity and magnetism, the generation of a magnetic field requires the core to be metallic, that is, electrically conducting in order for the geodynamo to function. (2) Studies of the abundance of materials throughout the cosmos, other than Fe, that would serve the dynamo function, fail to suggest any alternatives for Fe. Probably the principal consensus concerning the Earth's core pertains to acceptance of Fe as the primary element of the core. There is no consensus concerning the degree of purity of the core iron. Ferric alloys that would seem feasible at the extremes of temperature and pressure in the core have been proposed by some researchers. Other materials suggested have included iron sulfide (FeS), which is a good electrical conductor. The presence of iron oxides in the core has been suggested. Some investigators point to the composition of meteorites, iron and stony, as possibly indicative of the two core layers of Earth. Silicon-oxygen compounds make up rocks. As pointed out by Jeanloz, however, beyond this, there is no evidence for a more detailed analogy between the Earth's core and the nature of meteorites.

If the accuracy of core density estimates hold up, then it is evident that the percentage of lighter-than-iron elements cannot be extensive.

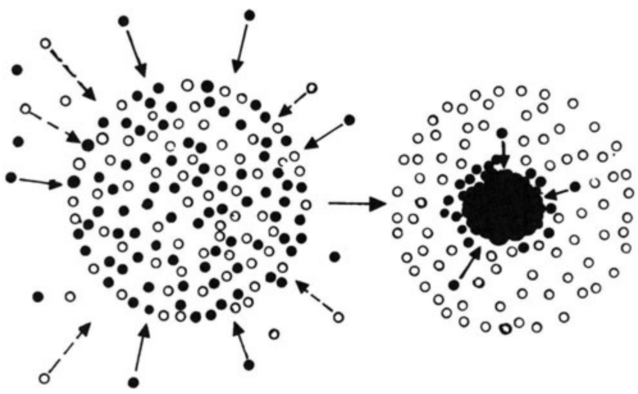
There has been much speculation pertaining to the driving force which propels the dynamo. Two distinct mechanisms have been proposed; one thermal, the other compositional. The *thermal concept* requires no difference in composition between the inner and outer cores, but it requires a source of energy. Radioactive isotopes, such as U-238 or K-40, which are present in the crust and the mantle, have been suggested. According to present theory, radioactive decay is the leading candidate as the dominant source of energy driving the flow in the core. John Verhoogen (University of California at Berkeley) has investigated another hypothesis, namely, that if the inner core is freezing out of the surrounding liquid, there could be sufficient heat from the latent heat of crystallization to power the geodynamo. There is also the "primordial heat" hypothesis. This is briefly explained in Fig. 9. Considerably more detail is given in the Jeanloz paper.

The composition concept for powering the dynamo is based upon the differences in densities of materials. No temperature differential is required to support this concept. The concept does require that the liquid of the outer core can separate into two phases (presumably solid and liquid) of significantly different composition, and thus significantly different densities. As concisely described by Jeanloz, a liquid with a composition different from that of a solid is exactly what would be expected for a partially frozen alloy under equilibrium conditions. This is why an alloy melts and freezes at slightly different temperatures. The presumption is that the solid inner and liquid outer parts of the core are at equilibrium and that accordingly they differ in composition. The presence of a seismic attenuating zone at the top of the inner core—possible seismological evidence for a crystal-liquid mush—tends to support this concept. Geophysicists are evaluating this model, but more seismological data and experiments at high pressure are needed. Regarding the laboratory production of high pressure, see article on **Diamond Anvil High Pressure Cell.**

In considering the two foregoing concepts regarding the powering of the core dynamo, it is interesting to note that a thermally driven geodynamo would require no composition difference; and a composition-driven geodynamo would require no temperature difference.

Explanation of Mantle Dynamics

As of the early 1990s, geophysicists are giving much attention to how the lower and upper mantle regions of the Earth's interior function and possibly interact with each other to produce the numerous tectonic (earthquakes) and volcanic (eruptions) phenomena that almost continuously affect some portion of the Earth's crustal (outer) surface. Surface effects are well known, and the fact that these effects arise from forces within the Earth's interior is self-evident. What remains poorly under-



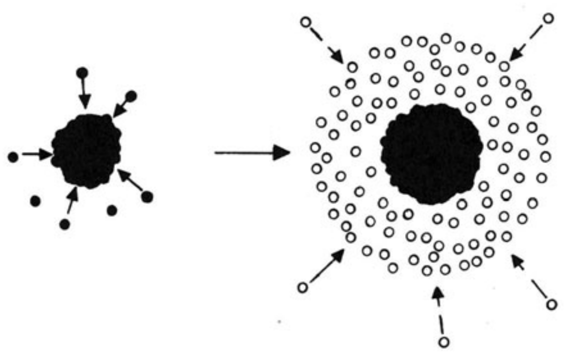


Fig. 9. In the *homogeneous accretion model* (top diagram), silicate and iron accumulated to form the complete planet as shown at the left. Sometime later, the core formed as the result of separation of the metal from the silicates. As the core formed, the iron sank to the center of the planet. Heat was generated by the release of gravitational energy.

In the *heterogeneous accretion model* (bottom diagram), it has been proposed that the core accumulated first as shown at left, after which the silicate mantle accreted around it. This is a sequence that could have taken place during or after condensation of solids out of the solar nebula, as governed by whether chemical, physical, or both processes were involved. In these models, accretion of the planet is proposed to result from infall of meteoritic bodies, as indicated at right.

There are numerous variations of these models, as explained in the Jeanloz reference. (*Diagram is after Jeanloz.*)

stood is how mantle regions produce such effects—that is, what is going on in the mantle areas? There are numerous hypotheses, some of which have been described in this article. At a meeting of the American Geophysical Union (held in Montreal, Canada, in May 1992), three schools of thought were described by groups of geophysicists, with differing findings and opinions. One group essentially claimed that materials and energy rather freely circulate between the upper and lower mantle. Another group stressed that there is little, if any, mixing of upper and lower mantle materials—that is, there is a barrier between the two mantles. Still another group suggested that mixing between the two mantles may occur sometimes and at varying locations. This latter group, in essence, accepted a combination of the other two viewpoints.

Understanding the Earth's interior machine is far more difficult than learning about its exterior—that is, scientists can look back at the Earth from spacecraft and probes. Attempts to get inside the Earth by way of sinking deep holes and drilling cores is difficult and costly. Experiments with materials in laboratories under high-temperature and high-pressure conditions have contributed knowledge, and computer models can be created from known and assumed statistics. Some evidence of this type is less than convincing, but progress is being made toward achieving answers, but at a comparatively slow pace.

For further clarification, the Kerr article (December 4, 1992), listed at the end of this article, is suggested reading. Also, more information can be found in the articles on **Earth Tectonics and Earthquakes** and on Volcano contained in this Encycloopedia.

The Geomagnetic Field

The North Magnetic Pole.* During the 16th century, mariners believed that somewhere in the North was a magentic mountain which was the source of attraction for compasses and an unfortunate object for any ships which strayed too close. It was not until 1600 that a better explanation was suggested. Sir William Gilbert, physician to Queen Elizabeth I, suggested that the Earth itself was a giant magnet and that the force which directs the compass originates inside the planet. Using a model of the Earth made from lodestone, Gilbert also showed that there should be two points where a magnetized needle would stand vertically—the North and the South Magnetic Poles.

This is basically the same definition used today. At the magnetic poles, the Earth's magnetic field is perpendicular to the Earth's surface. Consequently, the magnetic dip, or inclination, is 90 degrees. Since the magnetic field is vertical, there is no force in a horizontal direction to direct a compass needle—thus it will not point to any preferred direction. The magnetic declination, the angle between true geographic North and magnetic North, therefore, cannot be determined at the magnetic pole.

Where Is the North Magnetic Pole? Gilbert believed that the magnetic pole coincided with the geographic pole. Magnetic observations made by explorers in subsequent decades have shown that this is not true. By the early 19th century, it became established that the pole must be located somewhere in Arctic Canada.

In 1829, Sir John Ross set out to discover the Northwest Passage, but his ship became trapped in ice off the northeast coast of the Boothia Peninsula, where it was to remain for the next four years. Sir John's nephew, James Clark Ross, used the time to take magnetic observations along the Boothia coast. These observations convinced him that the pole was not far away. In the spring of 1831, he set out to reach it. On June 1, 1831, at Cape Adelaide on the west coast of Boothia, he measured a dip of 89°59′. For all practical purposes, he had reached the North Magnetic Pole.

The next attempt to reach the magnetic pole was made some 70 years later by the Norwegian explorer, Roald Amundsen. In 1903, Amundsen left Norway on his famous voyage through the Northwest Passage, which, in fact, was his secondary objective. His primary goal was to set up a temporary magnetic observatory in the Arctic and to determine the position of the magnetic pole.

A pole position was next determined by Canadian government scientsts shortly after World War II. Paul Serson and Jack Clark of the Dominion Observatory of Canada measured a dip of 89°56′ at Allen Lake on Prince of Wales Island. This, in conjunction with other observations in the vicinity, showed that the pole had moved some 250 kilometers northwest since Amundsen's observation.

Subsequent observations have been made by Dominion Observatory personnel in 1962, by personnel from the Canadian Department of Energy, Mines and Resources in 1973, and most recently in May 1984. This latest survey located the pole at 77.0°N, 102.3°W off the southeast tip of Lougheed Island and showed that the general northwesterly motion of the pole is continuing. See Fig. 10. During this century, the pole has moved an average of 10 km per year.

It is important to appreciate that when referring to the location of the pole, we are referring to an *average position*. The pole wanders daily in a roughly elliptical path around the average position, and may frequently be as much as 80 km away when magnetic conditions are disturbed.

Why Is the Pole Moving? If, as Gilbert believed, the Earth acts as a large magnet, the pole would not move, at least not as rapidly as it does. We now know that the cause of the Earth's magnetic field is much more complex and believe that it is produced by electrical currents that originate in the hot core of the planet.

In nature, processes are seldom simple. The flow of electric currents in the core is continually changing—so the magnetic field those currents produce also changes. Thus, at the surface of the Earth, both the strength and direction of the magnetic field will vary over the years. This gradual change is called the *secular variation* of the magnetic field.

*Prepared by L. R. Newitt and E. R. Niblett, Geological Survey of Canada, Geophysics Division, Ottawa, Ontario, Canada.

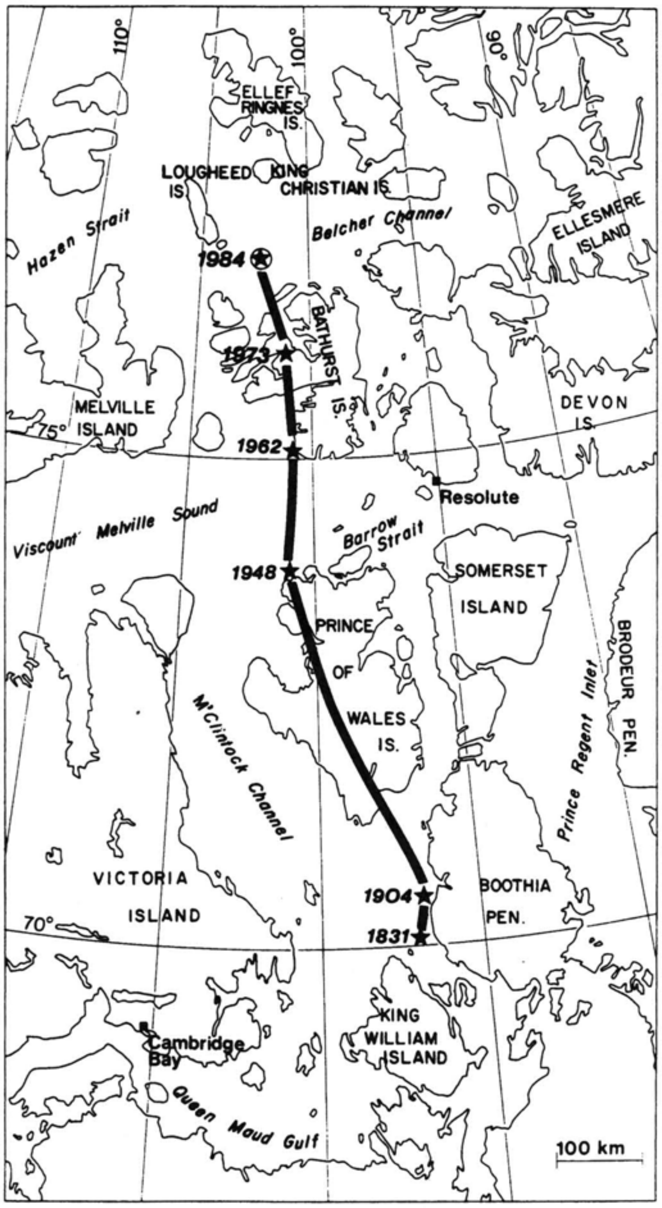


Fig. 10. Migration of the Earth's North Magnetic Pole in a northwesterly direction since first determined in 1831. In 1984, the pole was located at 77.0°N, 102.3°W off the southeast tip of Lougheed Island, Northwest Territories, Canada. The next official survey of the pole position will take place in 1994. Estimated position as of 1993 is 77.9°N, 103.2°W. (*Geological Survey of Canada, Geophysics Section, Ottawa, Ontario, Canada.*)

The position of the North Magnetic Pole is strongly influenced by the secular variation in its vicinity. For example, if the dip is 90° at any given point this year, that point will be the pole by definition. However, because of secular variation, the dip at that point will change to 89°55′ in about two years and it will no longer be the pole. However, at some nearby point, the dip will have increased to 90° and that point will then be the pole. In this manner, the pole slowly moves across the Arctic.

The more rapid daily motion of the pole around its average position has an entirely different cause. If we record the Earth's magnetic field continually, as is done at a magnetic observatory, we will see that it changes continuously during the course of a day—sometimes slowly, sometimes rapidly. The ultimate cause of these fluctuations is the sun. The sun constantly emits charged particles which upon encountering the Earth's magnetic field cause electric currents to be produced in the upper atmosphere. These electric currents disturb the magnetic field, resulting in a shift in the pole's position. The distance and speed of these displacements will, of course, depend on the disturbances in the magnetic field, but they occur constantly. When scientists try to determine the average position of the pole, they must average out all of these transient wanderings.

Importance of the Pole. The reasons that the magnetic pole causes so much interest have changed with time. For Ross, the search for the pole was a byproduct of scientific nationalism. For Amundsen, it offered a good excuse to sail through the Northwest Passage. Today, we are interested in the pole as a tool for magnetic cartography.

To understand why it is important, we must recall magnetic declination is the angle between *true north* and the direction in which a compass needle will point, that is, *magnetic north*. Declination changes from one location to the next. For example, it is 21°E in Vancouver, B.C., but in Boston it is 16°W. A knowledge of magnet declination is extremely important for navigational purposes. Therefore, maps showing the magnetic declination are revised and published every 5 years. These maps show contour lines along which the magnetic declination is equal. All these lines converge on the magnetic pole—so if the position assigned to the pole is wrong, the entire pattern of magnetic contours in the Arctic regions would be wrong.

Earth's Magnetotail. The solar wind is a continuous flow of subatomic particles, emanating from the sun, into space. These particles cannot easily penetrate the earth's magnetic field and in the region of the earth, the particles stretch out several millions of kilometers, forming an approximately cylindrical volume which E. W. Hones, Jr. (Los Alamos National Laboratory) likens to a *huge wind sock*. More officially, it is called the *earth's magnetotail*. Hones observes that the solar wind and the earth's magnetosphere form a vast electrical generator—one in which the interaction of magnetic fields and plasma (gaseous part of solar wind) converts the kinetic energy of the solar wind's motion into electricity. The solar wind also has been the target of considerable research in connection with space probe studies of other planets, notably of Venus, Mars, Jupiter, and Saturn.

Research in recent years has shown that the electricity generated through conversion of energy in the solar wind accounts for a number of interesting phenomena that for many years remained mysterious. Of central interest has been the beautiful displays of the auroras (See **Aurora and Airglow).** Another phenomenon is the presence of the Van Allen radiation belts which encircle the earth at altitudes of about 1000 and 6000 kilometers. (See **Van Allen Radiation Belts.**) Another phenomenon is the presence of plasmoids, which have been described as bodies of hot plasma held together by magnetic fields. Satellite research has shown that plasmoids behave like projectiles, achieving speeds in terms of millions of kilometers per hour.

The earth's magnetosphere is defined as that region around the earth to which the earth's magnetic field is confined, due to interaction between the solar wind and the geomagnetic field. On the sunlit side, the magnetosphere is approximately hemispherical, with a radius of about ten Earth radii under quiet conditions. It may be compressed to about six Earth radii by magnetic storms. On the side opposite the sunlit side, the magnetosphere extends in a "tail" of several hundred Earth radii.

Although detailed description is beyond the scope of this encyclopedia, three interacting forces exist in the magnetotail. These have been succinctly described by Hones as the Lorentz force, the **E**-cross-**B** drift, and the **J**-cross-**B** force. The Lorentz force causes electrons and protons to move in circles in opposite directions, thus "tying" the particles to magnetic field lines. If an electric field is imposed perpendicular to the magnetic field, the charged particles acquire a further motion known as the **E**-cross-**B** drift. The drift carries the centers of the particle's circular paths in a direction perpendicular to the directions of the field and thus plasma in the lobes of the magnetotail is driven toward the plasma sheet. Finally, plasma carrying an electric current that flows perpendicular to a magnetic field experiences a **J**-cross-**B** force. This force accelerates the plasma in a direction perpendicular to both the direction of the current and that of the magnetic field. This distorts the original field, bending the field lines in the direction opposite to that of the force. The **J**-cross-**B** force resisting the flow of solar-wind plasma distorts the earth's magnetic field, thus producing the magnetotail. These forces and interactions are excellently diagrammed and detailed in the Helms reference listed. See also **Magnetism;** and **Sun (The).**

Geomagnetic Anomalies. As previously mentioned, apparently the geomagnetic generator in the Earth's core is a mechanism (process) that does not operate with the precision and predictability associated with human-engineered electric power generators, as witnessed by the meandering North Magnetic Pole, by historically reported reversals of the geomagnetic field, and by specific situations such as the frequently mentioned East Coast Magnetic Anomaly. These anomalies (departures, abnormalities) do not all arise, of course, from some unsteadiness on the part of the core dynamo, but also from solar influences, as previously mentioned, and from underlying structures in the crust, lithosphere, and other layers of the Earth's interior, i.e., the mantles and possibly the outer core.

Scientists who have investigated magnetism recorded in a sequence of lava flows at Steens Mountain (Oregon) speculate that a reversal of the geomagnetic field occurred some 15 million years ago. Data indicate that the switch was from a reverse field to what is now considered the normal field (magnetic and geographical north being in the same direction). Investigators estimate that the North Magnetic Pole followed a convoluted path for nearly 15,000 years.

Still subject to some debate, scientists at the University of Paris reported in 1978 that the Earth's magnetic field had "shivered" or "jerked" in the late 1960s. There is agreement that the westward drift over Europe speeded up abruptly in 1969. As presently organized, data from magnetic observatories are considered inadequate. Most observatories are located in the Northern Hemisphere. Wide areas of the Pacific and Southern Hemisphere are not covered. Frequently data detected in Europe may not be evidenced in North America. No really profound explanations have been given for the 1969 instant; nor for another which some scientists believe occurred in 1912. Most geophysicists agree that a satellite or a series of satellites may be the only practical solutions to magnetic data problems.

Magnetotactic Bacteria. Several species of magnetotactic bacteria, as reported by Frankel, et al., have been observed in aquatic sediments of the Northern and Southern Hemispheres. Each bacterium contains magnetosomes consisting of enveloped, single-domain magnetite particles. The magnetosomes are often arranged in chains with a magnetic dipole moment parallel to the axis of motility sufficiently large that the cell is oriented along the geomagnetic field lines as it swims. Cells with North-seeking pole forward swim North along the magnetic field lines; the opposite occurs in cells with the South-seeking pole forward. Because of the inclination of the geomagnetic field, North-seeking cells migrate downward in the North Hemisphere and upward in the Southern Hemisphere. Magnetotactic bacteria are present in fresh water and marine sediments of Fortaleza, Brazil, situated close to the geomagnetic equator and, not surprisingly, are found roughly in equal numbers in samples taken.

Geodesy

Geodesy is principally concerned with the size and shape of the Earth and its gravitational field (from a practical and application standpoint rather than theoretical). Key geodetic parameters of the Earth are given in Table 2. The first recorded effort to estimate the circumference of the Earth dates back to Eratosthenes of Alexandria in the 3rd century B.C. Eratoshenes observed that the sun was overhead at Aswan at mid-summer because it shone directly down a well, but in Alexandria, it was 7.2° or $\frac{1}{50}$th of a circle away from the vertical. The distance between Aswan and Alexandria was calculated by estimating the time required by a camel to traverse the distance. The calculation was amazingly correct—within 1% of the currently accepted value. This was an example of space geodesy, i.e., using the sun as the reference object, a technique that surfaced again after the launching of the first artificial satellite, Sputnik, in 1957.

TABLE 2. GEODETIC PARAMETERS OF THE EARTH

Parameter	Standard Value	Current Estimate and Standard Deviation
Mean sidereal rotation rate. ω	$0.7292115085 \times 10^{-4}$ sec^{-1}	$0.729115085 \times 10^{-4}$ sec^{-1}
Equatorial gravity, γe	9.780490 m/sec^2	9.780306 ± .000013 m/sec^2
Equatorial radius, *a*	6378388 m	6378160 ± 15 m
Flattening, *f*	1/298.25 ± .03	1/298.25 ± .03

The Geoid. Because of the rotation of the Earth, its lack of absolute rigidity, crustial mass distribution, and tidal forces, the shape is not perfectly spherical, but approximates that of a triaxial ellipsoid. There is flattening at the poles (polar diameter is shorter than the equatorial). The actual figure of the Earth, which is irregular (not just at the poles) is referred to as *the geoid*. See Fig. 11. The largest departure of the actual geoid from the reference geoid (mapped) occurs in a depression of approximately 113 meters (371 feet) south of India. Off New Guinea, the highest hump occurs, about 81 meters (266 feet). There are two other significant humps, one located south of the British Isles and about 61 meters (200 feet) high; and the other hump located south of Madagascar and about 56 meters (184 feet) high.

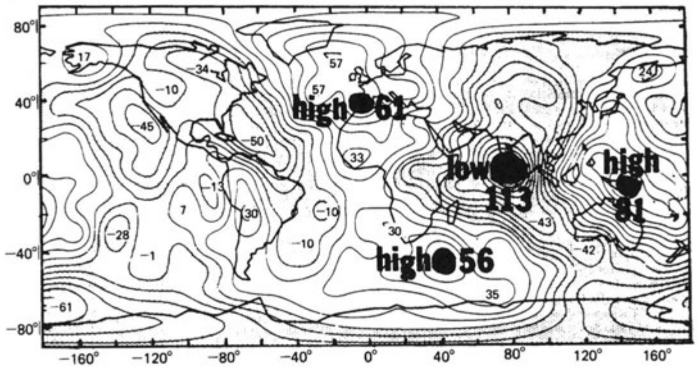


Fig. 11. In 1970, the Smithsonian Institute released what became known as the Smithsonian Standard Earth II. Preparation of the contours shown represented many years of work, using cameras and gravimetric surveys, as well as the first results of measurements by laser tracking of satellites. Over 100,000 photographic observations became part of the data bank. Calculations involved some 200,000 simultaneous equations, solving for about 400 unknowns. These included nearly 300 harmonic coefficients and the station coordinates. This map shows contours of the geoid at 10-meter (32.8-foot) intervals, which are relative to a reference spheroid of flattening 1 part in 198.25. Depressions are shaded areas. The three largest highs and the principal low are indicated. See test. (*After Smithsonian Institution map.*)

In constructing the dimensions of the geoid, one is not concerned with the regular physiographic features of the Earth—mountains, ocean basin, etc. The concern is the shape of the mean sea-level surface, continued under the land in a logical fashion. This surface is exclusively defined by measuring the variations of the Earth's gravitational attraction, both with latitude and longitude. To this must be added the acceleration produced by the earth's rotation. Bomford (1971) defines the geoid as "an equipotential surface of the Earth's gravitational potential plus the rotational potential." The geoid is the basic reference shape on which the Earth's topography (height above sea level) is superposed.

There is a dual relationship between the geoid and orbiting satellites. As pointed out by King-Hele (1976), one effect of the Earth's flattening on a satellite orbit is to make the plane of the orbit rotate about the Earth's axis in the direction opposite to the satellite's motion, while leaving the orbit inclined at the same angle to the equator. See Fig. 12. As further explained by King-Hele, the most helpful way of analyzing the shape of the geoid is to assume that if it made up of an infinite number of harmonics, the second harmonic defining the flattening, the third harmonic often being called pear-shaped, the fourth harmonic square-shaped, etc. See Fig. 13. Shapes of these kinds would be obtained if one sliced the Earth through the poles if only one particular

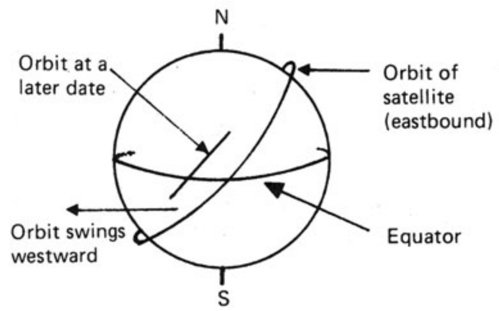


Fig. 12 Gravitational pull of the Earth's equatorial bulge causes orbital plane of an eastbound satellite to swing westward. (*After King-Hele.*)

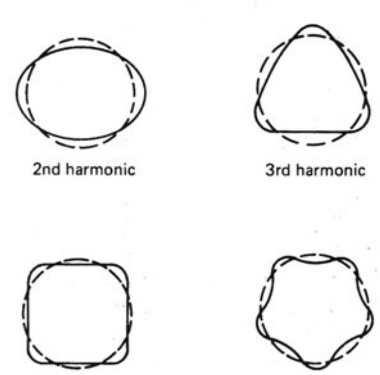


Fig. 13. Harmonics defining the characteristics of the geoid. There are many more harmonics. (*After King-Hele.*)

harmonic existed. But, in fact, all harmonics exist. Thus the procedure is one of calculating them separately and then putting them together to yield the final shape.

Laser tracking of satellite has been used for a number of years in making observations for determining the geoid, but these methods have been greatly improved since the mid-1970s. A satellite with corner reflectors was put in orbit in 1965 and followed by a number of like satellites for laser tracking. One method of investigation is to consider the satellite as points in the sky for geometrical triangulation; the other method analyzes the effect of the Earth's gravitational attraction on satellite orbits, thus determining how gravity changes over the Earth. The earlier satellites were in the *Beacon* and the *Lageos* series. The first of the *Magsat* (Magnetic Field Satellite), jointly sponsored by the National Aeronautics and Space Administration (NASA) and the U.S. Geological Survey was launched in October 1979. This was designed as a pole-crossing satellite with a range from 350 to 500 kilometers (217 to 311 miles) with an expected life of about 120 days.

One may ask, why is it so important to accurately determine the shape of the geoid? The effect of the gravitational pull of the Earth's equatorial bulge and other irregularities affect the orbit of an artificial satellite. But why were these humps and depressions important prior to the satellite age? For one thing, it has been determined that these humps and depressions tend to migrate with time and undoubtedly are related to migrations of the magnetic poles. (It has not been determined which is the cause and which is the effect.) With greater accuracies of measurement obtainable, as from Doppler radar and laser methods, satellite geodesy is making and will continue to impact all of the Earth sciences. This knowledge will assist in making crucial tests of theories about the Earth's interior, with some aspects of tectonic plate maps already being reflected in geoid mappings. Further, patterns of convection currents or density irregularities within the Earth will have to be meshed with the observed gravitational field. Combined with satellite altimeter measurements, gravitational field determinations will provide an accurate profile of the geoid surface, including earth tides and ocean tides. More knowledge of the polar motion (locus of the point where the earth's axis of rotation cuts the surface) most likely will also be obtained from this line of research. This knowledge may also assist in refining present theories of the Earth's magnetism and provide more plausible explanations of how the Earth's magnetic field is created in the first place.

The measured values of gravity depend on latitude because of the flattening and the variation of the centrifugal force from pole to equator. The normal value of gravity at the Earth's surface in centimeters per second per second is represented by

$$\gamma = 978.0516[1 + 0.005291 \sin^2 \phi - 0.0000057 \sin^2 2\phi + 0.0000 \cos^2 \phi \cos 2(\lambda + 6°)]$$

where ϕ and λ are latitude and longitude, respectively.

Gravity measurements have shown that in spite of large mass difference at the surface, the Earth is nearly in isostatic equilibrium. Various crustal blocks act as if they floated in a dense subcrustal material. The undulations of the geoid do not exceed 80 meters. Approximate dimensional figures for the Earth are: surface area, 510.1×10^6 square kilometers; volume 1.083×10^9 cubic kilometers; average density 5.517 grams per cubic centimeter; mass; 5.975×10^{27} grams; equatorial radius, 6.378 388 kilometers.

The deformation of the solid Earth by tidal forces form a specialty. The twice-daily occurring tides are observed by deflections of the vertical or variations of gravity. For the lunar tide, the variations amount to about 0.168 milligal; for the sun, up to 0.075 milligal. (One milligal equals 10 micrometers per second per second.) The maximal elevation of the geoid is 36 centimeters, the largest depression 18 centimeters, for the lunar effect; the total solar tide can reach 25 centimeters. The combined total at new and full moon is 79 centimeters.

There are five principal systems of measurement in geodesy:

(1) *Horizontal control* comprises the determination of the horizontal components of position—latitude and longitude—starting from fixed values for a certain point. It includes measurement of distances over the ground by metal tapes or by pulsing or modulating radio or light signals, and measurement of angles about a vertical axis by theodolites. Over the land, the relative horizontal position of points is obtained either by triangulation—a system of overlapping triangles with nearly all angles measured, but only occasional distances measured; or by traverse—a series of measured distances at measured angles with respect to each other; or by trilateration—a system of overlapping triangles with all sides measured. Much of the land area of the world is covered by triangulation, which gives the difference in latitude and longitude between points in the same network with a relative error of about 10^{-5}.

(2) *Vertical control* comprises the determination of heights, which is performed separately from horizontal control because of irregularities in atmosphere refraction. The most accurate method, leveling, measures successive differences of elevation on vertical staffs by horizontal lines of sight taken at intermediate points over short distances (less than 150 meters) balanced so as to minimize differential refraction effect. The datum to which vertical control refers is mean sea level as determined by tide gages. The accuracy is such that the error in difference of elevation between points on the same principal network should be a few tens of centimeters or less.

(3) *Geodetic astronomy* comprises the determination of the direction of the gravity vector and the direction of the north pole at a point on the ground. Astronomic longitude is the angle between the meridian of the gravity vector and the Greenwich meridian and is determined by measuring the time of intersection of a line of sight by a star. In these types of astronomic observation, several stars are normally observed which are selected so as to minimize error due to atmospheric refraction. Astronomic azimuth is determined by the measurement of the horizontal angle between a target and Polaris or other reference star.

(4) *Gravimetry* comprises the determination of intensity of gravitational acceleration. Most gravimetric observations are made differentially, by determining the change, with change in location, of the tension on a spring supporting a constant mass. These measurements

are connected through a system of reference stations to a few laboratory determinations of absolute acceleration of gravity. The relative accuracy of gravimetry is about ±0.001 centimeters per second squared on land and ±0.005 centimeters per second squared at sea. The principal difficulty in its geodetic application is irregular distribution of observations.

(5) *Satellite tracking* comprises the determination of the directions, ranges, or range rates of Earth satellites from ground fixed stations. These observations will be affected both by errors in positions of the station with respect to the Earth's center of mass and by perturbations of the orbit by the Earth's gravitational field; hence, in conjunction with a suitable dynamical theory for the orbit, they are used to determine the position of tracking stations and the variations of the gravitational field. To minimize refraction effect, directions are determined by photographs of the satellite against the background of fixed stars. Satellites also can be used as elevated targets by simultaneous observations from several ground stations.

The principal practical application of geodesy is to provide a distribution of accurately measured points to which to refer mapping, navigation aids, engineering surveys, geophysical surveys, and so on. The principal scientific interest in geodesy is the indication of the Earth's internal structure by the variations in the gravity field.

Meteorology and Aeronomy

These fields are concerned with the physical state and the motions of the atmosphere, which is divided into a number of layers. The lowest is the troposphere with an average thickness of 7 to 8 kilometers in polar regions and 13 kilometers in the equatorial zone. Temperatures decrease to the interface, called tropopause, with the next layer the stratosphere. At the tropopause, polar temperatures average around −55°C; in equatorial regions, −80°C. In the stratosphere, temperatures stay nearly isothermal with height and increase again above 25 kilometers. 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 atmospheric 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 sphere. 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 a criteria of age.

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 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. See also **Atmosphere (Earth); Meteorology**.

Geology and Mineralogy

Much of the scientific information pertaining to the Earth has been derived from the investigations of geologists and mineralogists over the years. Both geology and mineralogy are old, well established scientific fields, but generally are not included in delineations of geophysics. This illustrates the overlapping of fields of scientific interests that will continue as more and more knowledge of the earth and the cosmos is collected. The more established fields expand their spheres of interest, whereas some of the newer specializations encroach upon the older fields. An interesting, fine distinction between geophysics and geology is noted the "Glossary of Geology" (American Geological Institute). The definition of geology commences, "Geology—study of the planet Earth." The definition of geophysics commences, "Geophysics—study of the Earth as a planet." Mineralogy, of course, is the study of minerals, their formation, and occurrence. See also **Geology; Mineralogy.**

Geochemistry

Goldschmidt (1954) defined geochemistry as the study of the distribution and amounts of the chemical elements in minerals, ores, rocks, soils, water, and the atmosphere, and the study of the circulation of the elements in nature, on the basis of the properties of their atoms and ions; also, the study of the distribution and abundance of isotopes, including problems of nuclear frequency and stability in the universe. A more succinct definition would be that geochemistry is the study of the chemical constitution of the Earth and its chemical changes, either taking place now or having taken place.

Many inferences about the nature and composition of the different zones of the Earth are speculative to varying degrees because of the inaccessibility of the interior and because of the differentiated nature of the Earth. The overall composition may be deduced from spectroscopic evaluation of solar and stellar radiation, from nuclear chemical and astronomical theories of the origin of the elements and the evolution of the solar system, and from analytical study of the meteorites.

Oceanography

Several major articles in this encyclopedia are devoted to this important phase of the earth sciences. See **Ocean** and the articles which follow. Also, consult alphabetical index.

Selected Parameters and Characteristics of the Earth

Refer to following tables:

Table 3. Parameters of the Planet Earth
Table 4. Major Physical Features of the Planet Earth
Table 5. Officially Recorded Meteorological Extremes
Table 6. Major Floods and Tidal Waves
Table 7. Representative Major Earthquakes
Table 8. Representative Major Volcanoes and Eruptions
Table 9. Chemical Composition of Earth's Crust, Oceans, and Atmosphere

GLOBAL CHANGE: THE CONCEPT

Prolog. "Global change" is a phrase generally used to connote some form of degradation of the Earth's atmosphere, land surface, and hydrosphere which may result from human actions and disregard. Thus, depending upon what geobiological, geochemical, or geophysical consequence of the moment may be, global change will suggest numerous other phrases. Among these are air pollution, acid rain, smog, greenhouse effect, ozone layer depletion, water pollution, waste disposal, biodegradability, product recycling, environmental protection, population pressure, famine, species and biodiversity endangerment, deforestation, light pollution, noise pollution, toxic and nuclear wastes, fossil fuels, carcinogenicity, nuclear proliferation, and numerous other practices that do or that ultimately will affect the planet Earth and the life which it supports.

The term "global change" generally embraces those deleterious effects which arise from the life-styles and industrial pursuits of humankind, rather than the threatening effects which stem from purely natural causes that are not humanly manipulated. Some of these causes include leakage of petroleum into the oceans at subterranian depths, fouling of the atmosphere by ejecta from active volcanoes, and the introduction of massive volumes of carbon dioxide and other greenhouse gases from "inactive" volcanoes, not to mention large amounts of CO_2 created by the natural metabolisms of mammals, insects, and other living species.

TABLE 3. PARAMETERS OF THE PLANET EARTH

Geometric		
	Metric Units	English Units
Mass	5.9763×10^{27} grams (1/331950 mass of sun)	6 sextillion, 588 quintillion short tons
Volume	1083.1579×10^{9} km^3	259.8 bil cu mi
Surface Area	510.0501×10^{6} km^2	196,950,711 sq mi
Mean Radius	6370.949 km	3958.7 mi
Length of Equator	40,075.16 km	24,901.6 mi
Length of Meridian	40,008 km	24,860.0 mi
Distance from Sun	149,599,000 km	92,956.5 mi
Distance from Moon	356,410 to 406,697 km	221,463 to 252,710 mi
Length:		
Degree of Longitude at Equator*	111.324 km	69.173 mi
Degree of Latitude at Equator	110.551 km	68.693 mi
Degree of Latitude at Poles	111.669 km	69.388 mi
Magnetic North Pole (See Fig. 8)	77.0°N; 102.3°W (1987)	

*Varies as the cosine of the latitude.

Axis and Rotation	
Tilt of Earth's Axis away from Perpendicular to Orbit	23°27′
Period of Rotation on its Axis:	
Mean Solar Day	24 hours
Mean Sidereal Day	23 hours, 56 minutes, 4.091 seconds of mean solar time
Period of Revolution about the Sun	
Tropical Year	365 days, 5 hours, 48 minutes, 46 seconds (decreasing at rate of 0.530 second per century)

Variations in Rotation of the Earth

Secular—A slow secular increase in the length of the day (about 1 millisecond per century) is caused by tidal friction acting as a brake.

Irregular—Believed to be caused by turbulent motion in the core of the Earth, the speed of rotation may increase for a short period, such as 5 to 10 years, and then commence decreasing. During a century, the maximum difference from the mean in the length of a day is about 5 milliseconds. Since 1900, the accumulated *difference* (although a compensating effect) has amounted to approximately 40 seconds.

Periodic—There are seasonal variations with periods of 1 year and 6 months. The resulting cumulative effect is that each year the Earth is late about 30 milliseconds near June 1 and is ahead that same amount near October. It is believed that the semi-annual effect arises from tidal action of the sun, which distorts the shape of the Earth slightly. The maximum seasonal variation in length of the day is about 0.5 millisecond. This annual effect is probably due to the seasonal change in the wind patterns of the Northern and Southern Hemispheres.

Zonal Points and Parallels

		Latitude	
	North Pole	90°00′N	
Frigid Zone			
	Arctic Circle	66°30′N[a]	Northern
North Temperate Zone			Hemisphere
	Tropic of Cancer	23°27′N	
Torrid Zone	Equator	00°00′	
	Tropic of Capricorn	23°27′S	Southern
South Temperate Zone			Hemisphere
	Antarctic Circle	66°30′S[a]	
Frigid Zone			
	South Pole	90°00′S	

[a]Commonly defined as 23°30′ from the respective pole.

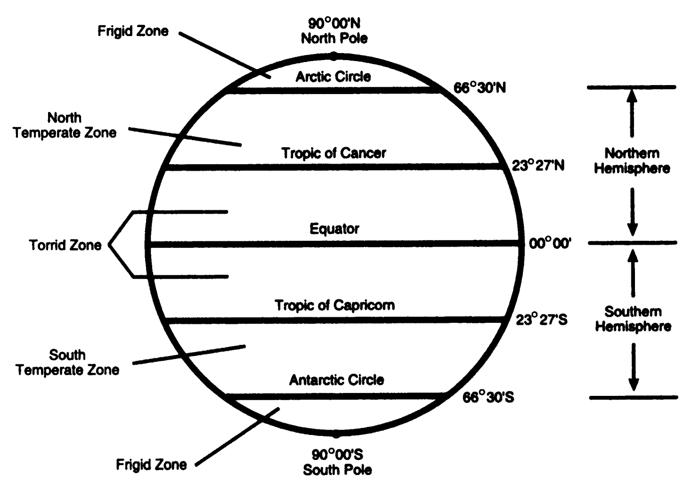


TABLE 4. MAJOR PHYSICAL FEATURES OF THE PLANET EARTH

Continents

Dimensions of Continents

Continent	Area		Dimensions			
			North to South		East to West	
	km^2	mi^2	km	mi	km	mi
Asia, including islands	44,750,000	17,276,909	8528	5300	9654	6000
Africa	30,280,000	11,688,728	8045	5000	7652	4700
North America, including islands	24,260,000	9,368,446	8528	5300	6436	4000
South America	18,050,000	6,970,760	7642	4750	4988	3100
Antarctica	13,380,000	5,165,000	—	—	—	—
Europe	10,220,000	3,947,441	3862	2400	6275	3900
Australia	8,530,000	3,294,866	3170	1970	3862	2400

Highest Continental Points

	Elevation above Sea Level	
Asia—Mount Everest, Nepal-Tibet*	8,847.7 m	(29,028 ft)
South America—Mount Aconcagua, Argentina	6,959.8 m	(22,834 ft)
North America—Mount McKinley (U.S.A.)	6,193.5 m	(20,320 ft)
Africa—Kibo (Kilimanjaro), Tanzania	5,894.8 m	(19,340 ft)
Europe—Mount El'brus (U.S.S.R.)	5,641.8 m	(18,510 ft)
Antarctica, Vinson Massif	4,138.9 m	(16,860 ft)
Australia, Mount Kosciusko	2,228 m	(7,310 ft)

Lowest Continental Points

	Depth below Sea Level	
Asia—Dead Sea, Israel-Jordan	395.9 m	(1,299 ft)
Africa—Lake Assal, Djibouti	156.1 m	(512 ft)
North America—Death Valley (U.S.A.)	86 m	(282 ft)
South America—Salinas Grandes, Peninsula Valdés, Argentina	39.9 m	(131 ft)
Europe—Caspian Sea (U.S.S.R.)	28 m	(92 ft)
Australia—Lake Eyre	15.9 m	(52 ft)

*Based upon Surveyor General of the Republic of India (1954); ±3 m (10 ft) because of snow. Concern has been expressed in recent years that another peak, K2, in the Himalayas may be slightly higher than Mt. Everest.

Economic Areas

Deserts	12.7×10^6 km^2	(4,904,593 mi^2)
Steppes	49.77×10^6 km^2	(19,217,465 mi^2)
Fertile Areas	87×10^{10} km^2	(33,588,038 mi^2)

Oceans

Ocean	Area		Average Depth	
	km^2	mi^2	meters	feet
Pacific	166,000,000	64,200,000	3940	12,925
Atlantic	86,000,000	33,400,000	3575	11,740
Indian	73,000,000	28,400,000	3872	12,598
Antarctic	36,000,000	13,900,000	—	—
Arctic	14,000,000	5,100,000	3840	3,407
ALL OCEANS	375,000,000	145,000,000	—	—

Greatest Ocean Depths

Location/Name	Meters	Depth Fathoms	Feet
Pacific Ocean			
Mariana Trench	10,924	5973	35,840
Tonga Trench	10,800	5906	35,433
Philippine Trench	10,057	5499	32,995
Kermadec Trench	10,047	5494	32,963
Bonin Trench	9994	5464	32,788
Kuril Trench	9750	5331	31,988
Izu Trench	9695	5301	31,808
New Britain Trench	8940	4888	29,331
Yap Trench	8527	4663	27,976
Japan Trench	8412	4600	27,599
Peru—Chile Trench	8064	4409	26,457
Palau Trench	8054	4404	26,424
Aleutian Trench	7679	4199	25,194
New Hebrides Trench	7570	4139	24,836
North Ryukyu Trench	7181	3927	23,560
Mid-America Trench	6662	3643	21,857

Location/Name	Meters	Depth Fathoms	Feet
Atlantic Ocean			
Puerto Rico Trench	8605	4705	28,232
South Sandwich Trench	8325	4552	27,313
Romanche Gap	7728	4226	25,354
Cayman Trench	7535	4120	24,721
Brazil Basin	6119	3346	20,076
Indian Ocean			
Java Trench	7125	3896	23,376
Ob' Trench	6874	3759	22,583
Diamantina Trench	6602	3610	21,660
Vema Trench	6402	3501	21,004
Agulhas Basin	6195	3387	20,325
Arctic Ocean			
Eruasia Basin	5450	2980	17,881
Mediterranean Sea			
Ionian Basin	5150	2816	16,896

Ranges of Spring Tides in Excess of 5 Meters (16.4 Feet)

Location	Tidal Range		Location	Tidal Range	
	Meters	Feet		Meters	Feet
Burntcoat Head, Nova Scotia			Dover, England	5.67	18.6
(Bay of Fundy)	14.49	47.5	Cherbourg, France	5.49	18.0
Rance Estuary, France	13.5	44.3	Antwerp, Belgium	5.43	17.8
Anchorage, Alaska	9.03	29.6	Rangoon, Burma	5.19	17.0
Liverpool, England	8.27	27.1	Juneau, Alaska	5.06	16.6
St. John, New Brunswick, Canada	7.20	23.6	Panama (Pacific Side)	5.01	16.4

Major Lakes

Name/Location	Area		Name/Location	Area	
	km^2	mi^2		km^2	mi^2
Caspian Sea (Asia/Europe)	371,002	143,244	Balkhash (Asia)	18,428	7115
Superior (N. America)	82,103	31,700	Ladoga (Europe)	17,703	6835
Victoria (Africa)	69,485	26,828	Chad (Africa)	16,317	6300
Aral Sea (Asia)	64,501	24,904	Maracaibo (S. America)	13,512	5217
Huron (N. America)	59,570	23,000	Onega (Europe)	9609	3710
Michigan (N. America)	57,757	22,300	Eyre (Australia)	9324	3600
Tanganyika (Africa)	32,893	12,700	Volta (Africa)	8485	3276
Baykal (Asia)	31,500	12,162	Titicaca (S. America)	8288	3200
Great Bear (N. America)	31,329	12,096	Nicaragua (N. America)	8029	3100
Malawi (Africa)	28,879	11,150	Athabasca (N. America)	7936	3064
Great Slave (N. America)	28,570	11,031	Reindeer (N. America)	6651	2568
Erie (N. America)	25,667	9,910	Turkana (Africa)	6405	2473
Winnipeg (N. America)	24,390	9,417	Issyk kul (Asia)	6100	2355
Ontario (N. America)	19,555	7,550	Torrens (Australia)	5776	2230

Major Rivers

Name	Flows Into	Length		Name	Flows Into	Length	
		km	mi			km	mi
Nile	Mediterranean	6693	4160	Yukon	Bering Sea	3184	1979
Amazon	Atlantic Ocean	6436	4000	Rio Grande	Gulf of Mexico	3033	1885
Chang Jiang	E. China Sea	6378	3964	Indus	Arabian Sea	2896	1800
Ob-Irtysh	Gulf of Ob	5409	3362	Brahmaputra	Bay of Bengal	2896	1800
Huang	Yellow Sea	4671	2903	Danube	Black Sea	2858	1776
Congo	Atlantic Ocean	4666	2900	Japura	Amazon River	2816	1750
Amur	Tatar Strait	4415	2744	Zambezi	Indian Ocean	2735	1700
Lena	Laptev Sea	4399	2734	Euphrates	Shatt-al-Arab	2735	1700
Missouri	Mississippi	4344	2700	Tocantins	Para River	2698	1677
Mackenzie	Arctic Ocean	4240	2635	Orinoco	Atlantic Ocean	2574	1600
Mekong	S. China Sea	4183	2600	Nelson	Hudson Bay	2574	1600
Niger	Gulf of Guinea	4167	2590	Paraguay	Parana River	2549	1584
Yenisey	Kara Sea	4092	2543	Amu	Aral Sea	2539	1578
Parana	Rio de la Plata	3998	2485	Ural	Caspian Sea	2534	1575
Mississippi	Gulf of Mexico	3778	2348	Ganges	Bay of Bengal	2510	1560
Murray-Darling	Indian Ocean	3717	2310	Salween	Andaman Sea	2414	1500
Volga	Caspian Sea	3530	2194	Arkansas	Mississippi	2348	1459
Purus	Amazon River	3379	2100	Colorado	Gulf of California	2333	1450
Madeira	Amazon River	3239	2013	Dnieper	Black Sea	2285	1420
Sao Francisco	Atlantic Ocean	3199	1988	Negro	Amazon River	2253	1400

TABLE 5. OFFICIALLY RECORDED METEOROLOGICAL EXTREMES

High Temperature

Temperature °C	°F	Location	Date
		Worldwide	
58	136.4	El Azizia, Libya (Africa)	September 13, 1922
57	134.6	Death Valley, California (United States)	July 10, 1913
54	129.2	Tirat Tsvi, Israel (Asia)	June 21, 1942
53	127.4	Cloncurry, Queensland (Australia)	January 16, 1889
50	122	Seville, Spain (Europe)	August 4, 1881
49	120.2	Rivadavia, Argentina (South America)	December 11, 1905
45	113	Midale, Saskatchewan (Canada)	July 5, 1937
		United States	
57	134.6	Death Valley, California	July 10, 1913
53	127.4	Parka, Arizona	July 7, 1905
50	122	Overton, Nevada	June 23, 1954
49	120.2	Alton, Kansas	July 24, 1936
49	120.2	Steele, North Dakota	July 6, 1936
49	120.2	Tishomingo, Oklahoma	July 26, 1943
49	120.2	Gannvalley, South Dakota	July 5, 1936
49	120.2	Seymour, Texas	August 12, 1936
49	120.2	Ozark, Arkansas	August 10, 1936
48	118.4	Pendleton, Oregon	August 10, 1898

Low Temperature

Temperature °C	°F	Location	Date
		Worldwide	
−89	−128.2	Vostok (Antarctica)	July 21, 1983
−68	−90.4	Verkkhoyansk/Oimekon (Asia)	February 6, 1933
−66	−86.8	Northice (Greenland)	January 9, 1954
−63	−81.4	Snag, Yukon (Canada)	February 3, 1947
−62	−79.6	Prospect Creek (Alaska)	January 23, 1971
−32	−27.4	Sarmiento, Argentina (South America)	January 1, 1907
−24	−11.2	Ifrane, Morocco (Africa)	February 11, 1935
−22	−7.6	Charlotte Pass (Australia)	July 22, 1947
		United States	
−62	−79.6	Prospect Creek, Alaska	January 23, 1971
−57	−70.6	Rogers Pass, Montana	January 20, 1954
−53	−63.4	Moran, Wyoming	February 9, 1933
−51	−59.8	Maybell, Colorado	January 1, 1979
−51	−59.8	Island Park, Idaho	January 18, 1943
−51	−59.8	Pokegama, Minnesota	February 16, 1903
−51	−59.8	Parshall, North Dakota	February 15, 1936
−51	−59.8	Strawberry, Utah	January 5, 1913
−50	−58	McIntosh, South Dakota	February 17, 1936
−48	−54.4	Seneca, Oregon	February 10, 1933
−48	−54.4	Danbury, Wisconsin	January 24, 1922
−47	−52.6	Old Forge, New York	February 18, 1979
−46	−50.8	Vanderbilt, Michigan	February 9, 1934
−46	−50.8	San Jacinto, Nevada	January 8, 1937
−46	−50.8	Gavilan, New Mexico	February 1, 1951
−46	−50.8	Bloomfield, Vermont	December 30, 1933
−44	−47.2	Washta, Iowa	January 12, 1912
−44	−47.2	Van Buren, Maine	January 19, 1925
−44	−47.2	Camp Clarke, Nebraska	February 12, 1899
−44	−47.2	Winthrop, Washington	December 30, 1968
−43	−45.4	Boca, California	January 20, 1937
−43	−45.4	Pittsburg, New Hampshire	January 28, 1925
−41	−42	Smethport, Pennsylvania	January 5, 1904
−40	−40	Hawley Lake, Arizona	January 7, 1971
−40	−40	Lebanon, Kansas	February 13, 1905
−40	−40	Oakland, Maryland	January 13, 1912
−40	−40	Warsaw, Missouri	February 13, 1905

TABLE 5. OFFICIALLY RECORDED METEOROLOGICAL EXTREMES *(continued)*

Precipitation				
	Amount			
Time Span	Cm	In	Location	Date
Rain				
1 Minute	3.1	1.2	Unionville, Maryland (U.S.)	July 04, 1956
20 Minutes	20.5	8.1	Curtea-de-Arges (Romania)	July 07, 1889
42 Minutes	30.5	12.0	Holt, Missouri (U.S.)	June 22, 1947
$\frac{1}{2}$ Day (12 hr)	114	44.9	Foc-Foc (Réunion)	January 07–08, 1966
1 Day (24 hr)	182.5	71.9	Foc-Foc (Réunion)	January 07–08, 1966
1 Day (24 hr)	125	49.2	Paishih (Taiwan)	September 10–11, 1963
1 Day (24 hr)	114	44.9	Bellenden Ker (Australia)	January 04, 1979
1 Day (24 hr)	109	42.9	Alvin, Texas (U.S.)	July 25–26, 1979
1 Day (24 hr)	49	19.3	British Columbia (Canada)	October 06, 1967
5 Days	395	155.5	Commerson (Réunion)	January 23–28, 1980
1 Month	930	366.1	Cherrapunji (India)	July 1861
1 Year	2647	1042.1	Cherrapunji (India)	August 1860–July 1861
1 Year	1878	739.4	Kukui, Maui, Hawaii (U.S.)	December 1981–November 1982
Snow				
19 Hours	173	68.1	Bessans (France)	April 05, 1969
1 Day (24 hr)	193	76	Silver Lake, Colorado (U.S.)	April 15, 1921
1 Day (24 hr)	158	61	Thompson Pass, Alaska (U.S.)	December 29, 1955
1 Storm (6 days)	480	189	Mt. Shasta, Ski Bowl, California (U.S.)	February 13, 1959[a]
1 Storm (7 days)	446	175.6	Thompson Pass, Alaska (U.S.)	December 26, 1955[a]
1 Month	991	390.2	Tamarack, California (U.S.)	January 1911
1 Season (2–5 mos)	2850	1122	Paradise Ranger St'n., Washington (U.S.)	1971–1972
1 Season (2–5 mos)	2475	974.5	Thompson Pass, Alaska (U.S.)	1952–1953
1 Season (2–5 mos)	2447	963.4	Revelstoke Mt., British Columbia (Canada)	1971–1972

[a]Commencement of storm.
Sources: U.S. Army Corps of Engineers and other official agencies.

TABLE 6. MAJOR FLOODS AND TIDAL WAVES

Year	Location	Fatalities	Year	Location	Fatalities
1228	Holland	100,000	1972	Man, West Virginia (U.S.)	118
1642	China	300,000	1972	Eastern Seaboard (U.S.)	129
1887	Huang He River, China	900,000	1972	Rapid City, South Dakota (U.S.)	237
1889	Johnstown, Pennsylvania (U.S.)	2,200	1974	Monty-Long, Bangladesh	2,500
1896	Sanriku, Japan	27,000	1976	Loveland, Colorado (U.S.)	139
1900	Galveston, Texas (U.S.)	5,000	1977	Andhra Pradesh, India	10,000
1911	Chang Jiang River, China	100,000	1981	Northern China	550
1931	Huang He River, China	3,700,000	1981	Sichuan, Hebei Prov., China	1,300
1939	Northern China	200,000	1982	Northern Peru (Lima)	600
1953	Netherlands/N.W. Europe	1,800	1982	Guangdong, China	430
1959	Frejus, France (dam collapse)	112	1982	Southern Connecticut (U.S.)	12
1960	S. California/Arizona (U.S.)	26	1982	El Salvador/Guatemala	1,300
1960	Bangladesh	4,000	1982	Illinois/Missouri/Arkansas (U.S.)	22
1960	Agadir, Morocco	11,000	1983	California Coast (U.S.)	13
1962	Peru (volcanic mudslide)	3,000	1983	Southeastern U.S.	15
1963	Italy (Vaiont Dam collapse)	2,000	1984	Tulsa, Oklahoma	13
1964	Alaska/U.S. West Coast	117	1984	South Korea	200
1970	East Pakistan	400,000	1985	Northern Italy (dam collapse)	361
1971	Orissa State, India	10,000	1988	Bangladesh	1,300

TABLE 7. REPRESENTATIVE MAJOR EARTHQUAKES

Year	Location	Estimated Fatalities	Estimated Magnitude
526	Antioch, Syria	250,000	
856	Corinth, Greece	45,000	
1057	Chihli, China	25,000	
1290	Chihli, China	100,000	
1293	Kamakura, Japan	30,000	
1556	Shaanxi, China	830,000	
1667	Shemaka, Caucasia	80,000	
1693	Catania, Italy	60,000	
1730	Hokkaido, Japan	137,000	
1737	Calcutta, India	300,000	
1755	Lisbon, Portugal	60,000	8.7
1755	Northern Persia (Iran)	40,000	
1783	Calabria, Italy	30,000	
1797	Quito, Ecuador	41,000	
1811–1812	New Madrid, Missouri (U.S.)	?	8.1–8.3
1828	Echigo, Japan	30,000	
1868	Peru/Ecuador	40,000	
1896	Japan (seawave)	27,000	
1906	San Francisco, California (U.S.)	500	6–7
1915	Avezzano, Italy	30,000	7.5
1920	Gansu, China	100,000	8.6
1923	Yokohama, Japan	200,000	8.3
1927	Nan-Shan, China	200,000	8.3
1932	Gansu, China	70,000	7.6
1935	Quetta, India	50,000	7.5
1939	Erzincan, Turkey	30,000	7.9
1939	Chillan, Chile	28,000	8.3
1964	Near Anchorage, Alaska (U.S.)	117	
1970	Peru	67,000	
1972	Iran	5,000	
1972	Nicaragua (Managua)	6,000	
1976	Guatemala	23,000	
1976	Tanshan, China	242,000	
1977	Bucharest, Rumania	1,540	
1978	Eastern Iran	25,000	
1980	Italy (Naples Region)	2,735	7.2
1980	Algeria (Northwest)	4,500	7.5
1983	Colombia (Southern)	250	5.5
1983	Turkey (Eastern)	1,300	7.1
1985	Chile	146	7.8
1985	Mexico City and Coastal States	25,000	8.1
1987	Whittier Narrows, California	0	5.9
1988	Armenia (Most damage caused by after shock of 5.8)	25,000	6.9
1989	Loma Prieta (Near San Francisco)	67	7.1
1990	Iran (Northwest near Caspian Sea)	110,000	7.7
1990	Luzon Island (Philippines)	1,600	7.7
1991	Afghanistan (Northern "Hindu Kush" Region)	400	6.8
1992	Landers, California	0	

NOTE: What constitutes a major earthquake? The number of human fatalities usually is a major criterion, but does not correlate necessarily with magnitude because some very strong earthquakes occur in sparsely populated areas, not in heavy population centers. Often, there is no correlation between lives lost and scientific interest. Property damage frequently is better correlated with building and housing construction than with the magnitude of an earthquake. This also pertains to lives lost because the majority of earthquake deaths result from falling construction materials that kill and injure (and even bury and suffocate) victims.

Society's concern with pollution dates back at least to the period of the industrial revolution which commenced in England in the mid-1800s, when steam-powered machinery was introduced, serving at least as a partial replacement of horse and human muscle power. Abhorrence of human disregard for the Earth's clean air was well portrayed by Charles Dickens in his novel, *Hard Times*.[1] Later, other writers (Sandburg, Munford, et al.) wrote of the veiled threats of industrialization, but mainly from a societal viewpoint rather than in terms of the direct effects, difficulty reversible, that were injuring the well being of the planet per se. Many years passed before a few forward-looking scientists warned of the long-term effects of Earth abuse. The scientific community, however, acted slowly. Within the last score of years, scientists have addressed a number of factors that are determining global change. In parallel, numerous advocacy groups were formed, each with specific agendas—"Save the Air," "Save the Fishes," "Save the Forests," and the like. Although such groups contributed to awareness, the tendency was that of developing a very mixed menu of objectives, seriously lacking any coordination.

Thus, until very recently, the problems of global change have been addressed in a partial, fragmented manner. That is why the Global Change Research Program, proposed by the National Aeronautics and Space Administration (U.S.) and announced in 1992, is a welcome first step toward conducting a concerted effort in the field.[2]

The NASA Global Change Research Program (GCRP). The overall objective of the NASA plan is that of gaining predictive understanding of the interactive physical, geological, chemical, biological, and social processes that regulate the total Earth system. Participants of the GRCP include NASA, the National Oceanographic and Atmospheric Administration (NOAA), the U.S. Departments of Energy, Defense, Agriculture, and the Interior, the Environmental Protection Agency, the National Science Foundation, and the Smithsonian Institution. Agencies from advanced foreign countries also will participate in some of the projects. Such participants include the World Meteorological Organization, the United Nations Environmental Program, the Inter-governmental Oceanographic Commission, and complementary Canadian, European, and Japanese Earth observing missions.

In total, the program is referred to as "Mission to Planet Earth." Some details of the program delineated in Fig. 14 and 15, include: extensive use of polar, geosynchronous, and other unmanned satellites equipped with a variety of the most advanced sensors, such as the cryogenic limb array Etalon spectrometer, which operates in the 3.5–12.7 microwave spectrum for determining thermal emissions from atmospheric water vapor, methane, ozone, nitrogen oxides, and chlorofluorocarbons; the advanced stratospheric and mesospheric sounder which operates in the 4.6–16.6 micrometer bands for determining nitrogen compounds; the halogen occultation instrument which operates at the 2.43–10.25 micrometer range for measuring vertical distribution of hydrofluoric and hydrochloric acids, methane, ozone, water vapor, and nitrogen compounds; the high-resolution Doppler imager, which observes emission lines of neutral and ionized atomic oxygen in the visible and near-infrared regions at altitudes above 15 km (9.3 mi); and a wind imaging interferometer that uses a high-resolution Michelson interferometer and detector array for measuring Doppler shifts in the emission lines of neutral and ionized atomic oxygen, two lines in the OH molecule, and a molecular oxygen line.

Further details of the GCRP are given in other articles throughout this Encyclopedia. In particular, reference is made to **Satellites (Scientific and Reconnaissance)** and **Atmosphere (Earth).** Also check Alphabetical Index.

Additional Reading

Airu, B.: "Mission to Planet Earth," *Sensors*, 48 (April 1992).

Allan, D. W., and M. A. Weiss: "Around-the-World Relativistic Sagnac Experiment," *Science*, 228, 69–70 (1985).

[1]Dickens, an early social reformer, observed, "(Coketown) was a town of red brick, or of brick that would have been red if the smoke and ashes had allowed it; but as matters stood it was a town of unnatural red and black like the painted face of a savage. It was a town of machinery and tall chimneys, out of which interminable serpents of smoke trailed themselves for ever and ever, and never got uncoiled. It had a black canal in it, and a river that ran purple with ill-smelling dye, and vast piles of buildings full of windows where there was a rattling and a trembling all day long, and where the piston of the steam engine worked monotonously up and down like the head of an elephant in a state of melancholy madness."

[2]It is only fair, however, to report that a lot of progress, although not well coordinated, has been made to ameliorate several causes of global change in recent years. One needs only to compare the improved conditions in several of the advanced industrial countries with the photographs of damage caused by careless, uncontrolled polluters in the former Soviet block.

TABLE 8. REPRESENTATIVE MAJOR VOLCANOES AND ERUPTIONS

Eruptions

7000 B.C.	Mazama (now Crater Lake, California)
79 A.D.	Mt. Vesuvius, Italy (also 1139, 1631, 1779, 1793, 1872, 1906, 1944)
1815	Tambora, Indonesia (also 1913)
1883	Krakatau, Netherlands East Indies
1902	Santa Maria, Quatemala
1902	Mt. Pelée, Martinique
1912	Katmai, Alaska
1914/1917	Lassen Peak, California
1964	Mt. Agung, Bali
1964	Huascaran, Peru
1968	Mt. Batur, Bali
1976	St. Augustine, Alaska (also 1812, 1883, 1902, 1935, 1963–1964)
1976	White Island Volcano, New Zealand
1978	Poas, Costa Rica
1978	Etna, Italy (over 70 eruptions in 2000 years)
1978	Arboukaba, Djibouti
1979	Fuego, Guatemala
1979	Soufrière, St. Vincent
1979	Karkar, Papua New Guinea
1979	Mt. Sinila, Java
1980	Mt. St. Helens, Washington (over 20 eruptions since 1900 B.C.)
1983	El Chichón, Mexico
1985	Nevado del Ruiz, Colombia
1991	Mt. Unzen, Japan
1991	Mt. Pinatubo, Philippines

Major Active and Inactive Volcanoes (Abridged List)

Name of Volcano	Location	Elevation Meters	Elevation Feet
Asia—Oceania			
Klyuchevskaya (1974)	Russia	4,850	15,913
Kerintiji (1968)	Sumatra	3,805	12,484
Fujiyama	Japan	3,776	12,390
Rindjani (1966)	Indonesia	3,726	12,225
Semeru (1976)	Java	3,676	12,061
Ichinskaya	Russia	3,631	11,913
Koryakskaya (1957)	Russia	3,528	11,575
Sundoro (1906)	Java	3,150	10,335
Agung (1964)	Bali	3,142	10,309
Mayon (1978)	Philippines	2,990	9,810
Apo	Philippines	2,953	9,689
Merapi (1976)	Java	2,911	9,551
Tambora (1913)	Indonesia	2,851	9,354
Ruapehu (1975)	New Zealand	2,796	9,174
Balbi	Solomon Islands	2,593	8,508
Asama (1973)	Japan	2,542	8,340
Niigata Yakeyama (1974)	Japan	2,400	7,874
Alaid (1972)	Kuril Islands	2,339	7,674
Ulawun (1973)	New Britain	2,248	7,376
Azuma (1978)	Japan	2,024	6,641
Nasu (1977)	Japan	1,917	6,290
Lamington (1952)	Papua New Guinea	1,830	6,004
Tiatia (1973)	Kuril Islands	1,822	5,978
Batur (1968)	Bali	1,717	5,633
Bagana (1960)	Solomon Islands	1,702	5,584
Keli Mutu (1968)	Indonesia	1,640	5,301
Lokon-Empung (1970)	Celebes	1,579	5,181
Lopevi (1960)	New Hebrides	1,447	4,748
Dukono (1971)	Indonesia	1,087	3,566
Suwanosezima (1979)	Japan	932	3,058
Usu (1978)	Japan	725	2,379
White Island (1979)	New Zealand	321	1,053
Taal (1977)	Philippines	300	984
North America			
Citlaltepec	Mexico	5,676	18,623
Popocatepetl (1920)	Mexico	5,452	17,880
Rainier	Washington (U.S.)	4,395	14,420
Wrangell	Alaska (U.S.)	4,320	14,174
Colima (1975)	Mexico	3,960	12,993
Spurr (1953)	Alaska (U.S.)	3,375	11,073
Baker	Washington (U.S.)	3,316	10,880
Lassen	California (U.S.)	3,186	10,453
Paricutin (1952)	Mexico	3,170	10,401
Shishaldin (1979)	Aleutians (U.S.)	2,858	9,377
St. Helens (1980)	Washington (U.S.)	2,549*	8,364
Katmai (1931)	Alaska (U.S.)	2,285	7,497
Trident (1963)	Alaska (U.S.)	2,070	6,792
Martin (1912)	Alaska (U.S.)	1,830	6,004
Cleveland (1951)	Aleutians (U.S.)	1,730	5,676
Westdahl (1978)	Aleutians (U.S.)	1,532	5,026
Augustine (1976)	Alaska (U.S.)	1,210	3,970
Seguam (1977)	Alaska (U.S.)	1,050	3,445
South America			
Guallatiri (1960)	Chile	6,060	19,883
Cotopaxi (1975)	Ecuador	5,897	19,348
El Misti	Peru	5,825	19,112
Tolima (1943)	Colombia	5,525	18,128
Purace (1977)	Colombia	4,600	15,093
Hudson (1973)	Chile	2,600	8,531
Fernandia (1977)	Galapagos Islands	1,546	5,072
Alcedo (1954)	Galapagos Islands	1,127	3,698
Antarctica			
Erebus (1975)	Ross Island	3,743	12,281
Big Ben (1960)	Heard Island	2,745	9,006
Melbourne	Victoria Island	2,590	8,498
Damley (1956)	South Sandwich Islands	1,100	3,609
Deception Island (1970)	South Shetland Islands	602	1,975
Mid-Pacific			
Mauna Kea	Hawaii (U.S.)	4,206	13,800
Mauna Loa (1978)	Hawaii (U.S.)	4,170	13,682
Haleakala	Hawaii (U.S.)	3,065	10,056
Kilauea (1992)	Hawaii (U.S.)	1,222	4,009
Central America			
Tajumulco	Guatemala	4,220	14,502
Tacana	Guatemala	4,092	13,426
Acatenango (1972)	Guatemala	3,976	13,045
Santa Maria (Santiaguito) (1979)	Guatemala	3,772	12,376
Fuego (1979)	Guatemala	3,736	12,258
Irazu (1967)	Costa Rica	3,432	12,245
Poas (1978)	Costa Rica	2,704	8,872
Pacaya (1978)	Guatemala	2,552	8,373
San Miguel (1976)	El Salvador	2,130	6,989
Izaico (1966)	El Salvador	1,965	6,447
Rincon de la Vieja (1968)	Costa Rica	1,806	5,925
El Viejo (San Cristobal)	Nicaragua	1,745	5,725
Arenal (1978)	Costa Rica	1,552	5,092
La Soufrière	Guadeloupe	1,467	4,813
Pelée	Martinique	1,397	4,584
Soufrière (1979)	St. Vincent	1,178	3,865
Masaya (1978)	Nicaragua	635	2,083
Europe			
Etna (1978)	Italy	3,290	10,794
Vesuvius (1944)	Italy	1,281	4,203
Stromboli (1975)	Italy	926	3,038

TABLE 8. REPRESENTATIVE MAJOR VOLCANOES AND ERUPTIONS *(continued)*

Name of Volcano	Location	Elevation Meters	Elevation Feet	Name of Volcano	Location	Elevation Meters	Elevation Feet
Vulcano	Italy	500	1,641	Teide (Tenerife) (1909)	Canary Islands	3,713	12,182
Santorini (1950)	Greece	130	427	Nyirangongo (1977)	Zaire	3,465	11,369
				Nyamuragira (1977)	Zaire	3,056	10,027
Mid-Atlantic Range				Ol Doinyo Lengai (1960)	Tanzania	2,886	9,469
Beerenberg (1970)	Jan Mayen Island	2,277	7,471	Fogo (1951)	Cape Verde Island	2,829	9,282
Tristan da Cunha (1962)	Tristan da Cunha Island	2,060	6,759	Piton de la Fournaise (1977)	Reunion	2,631	8,632
Askja (1961)	Iceland	1,510	4,954	Palma (1971)	Canary Islands	2,423	7,950
Leirhnukur (1975)	Iceland	650	2,133	Karthala (1977)	Comoro Islands	2,361	7,746
Krafla (1978)	Iceland	650	2,133	Erta-Ale (1973)	Ethiopia	615	2,008
Helgatell (1973)	Iceland	226	742				
Africa							
Kilimanjaro	Tanzania	5,895	19,341				
Cameroon	Cameroons	4,070	13,354				

NOTE: Volcanoes listed with no dates within parentheses have been inactive for several decades.
*Prior to May 18, 1980 eruption, elevation was 2,950 meters (9,677 feet).

TABLE 9. CHEMICAL COMPOSITION OF EARTH'S CRUST, OCEANS, AND ATMOSPHERE

Average Density	5.517 grams/cm^3
Estimated Density of Mantle	3–6 grams/cm^3
Estimated Density of Core	10–17 grams/cm^3

Occurrence of Elements in Earth's Crust

Element	%	Element	%	Element	%
Oxygen	47.	Sodium	2.5	Sulfur	0.1
Silicon	28.0	Potassium	2.5	Nickle	0.02
Aluminum	8.0	Titanium	0.4	Copper	0.002
Iron	4.5	Hydrogen	0.2	Lead and Zinc	0.001
Calcium	3.5	Carbon	0.2	Tin	0.00001
Magnesium	2.5	Phosphorus	0.1	Silver	0.000001
				All others	<0.48

Average Percentage of Metals in Igneous Rocks

Metal	%	Metal	%	Metal	%
Silicon	27.72	Copper	0.010	Niobium and Tantalum	0.003
Aluminum	8.13	Tungsten	0.005	Hafnium	0.003
Iron	5.01	Lithium	0.004	Thorium	0.002
Calcium	3.63	Zinc	0.004	Lead	<0.002
Sodium	2.85	Chromium	0.037	Cobalt	0.001
Potassium	2.60	Zirconium	0.026	Beryllium	0.001
Titanium	0.63	Nickel	0.020	Strontium	<0.001
Manganese	0.10	Vanadium	0.017	Uranium	<0.001
Barium	0.05	Rare earths	0.015		

Average Salt Content of Ocean Water 3.5% (wt)

	(Of Total Salts)
Sodium chloride $NaCl$	77.76%
Magnesium chloride $MgCl_2$	10.88
Magnesium sulfate $MgSO_4$	4.74
Calcium sulfate $CaSO_4$	3.60
Potassium sulfate K_2SO_4	2.46
Magnesium bromide $MgBr_2$	0.22
Calcium carbonate $CaCO_3$	0.34

Composition of the Atmosphere

Element	%(vol)
Nitrogen	78.03
Oxygen	20.99
Argon	<0.94
Carbon dioxide	0.03
Hydrogen	0.01
Neon	0.00123
Helium	0.004
Krypton	0.00005
Xenon	0.000006

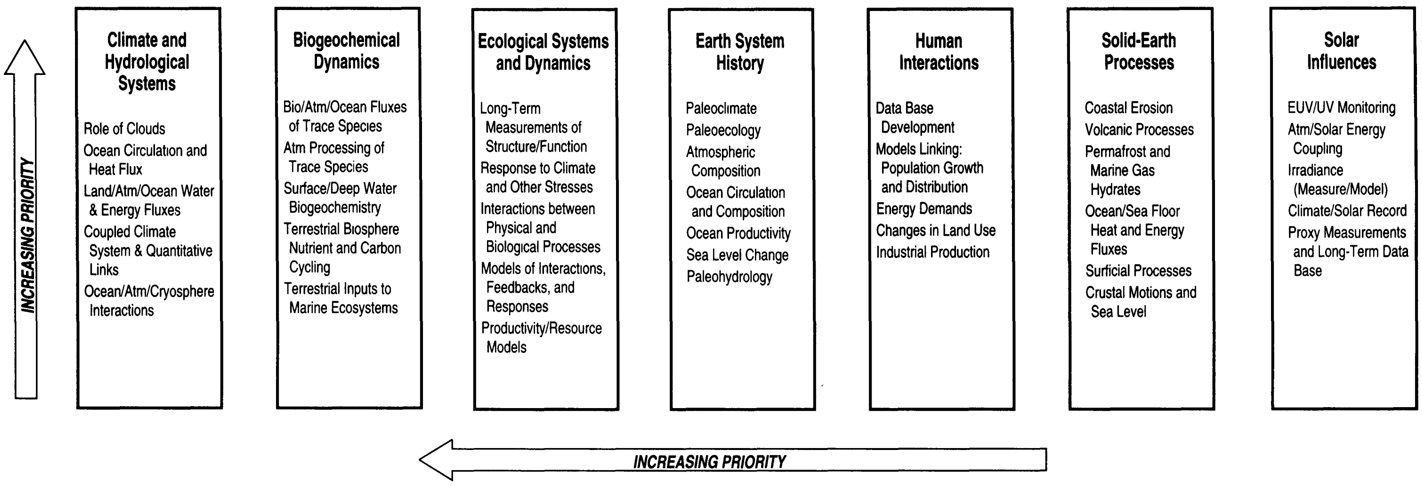


Fig. 14. Seven major areas of interest to be investigated as part of the Global Change Research Program. (*National Aeronautics and Space Administration.*)

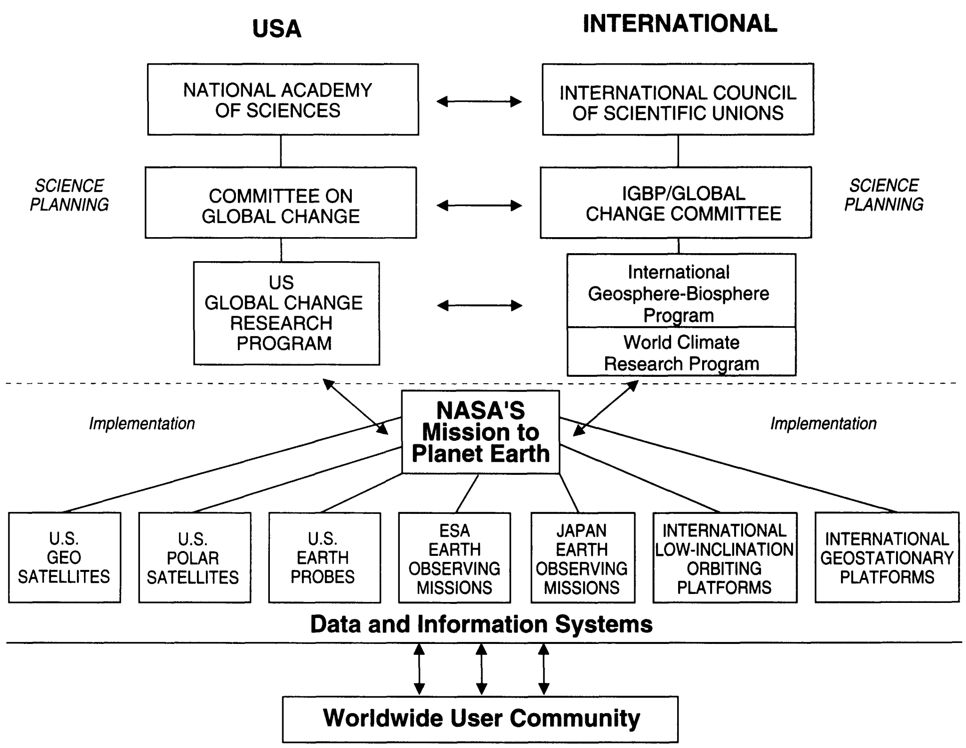


Fig. 15. General organization of the Global Change Research Program. (*National Aeronautics and Space Administration.*)

Alsop, L. E., and M. Talwani: "The East Coast Magnetic Anomaly," *Science*, **226**, 1189–1191 (1984).
Appenzeller, T.: "Roving Stones," *Sci. Amer.*, 19 (February 1990).
Bell, D. R., and G. R. Rossman: "Water in the Earth's Mantle: The Role of Nominally Anhydrous Minerals," *Science*, 1391 (March 13, 1992).
Bloxham, J., and D. Gubbins: "The Evolution of the Earth's Magnetic Field," *Sci. Amer.*, 68 (December 1989).
Bogorodsky, V. V., Bentley, C. R., and P. E. Gudmandsen: "Radioglaciology," Reidel (Kluwer, Hingham, Massachusetts), 1985.
Chorley, R. J., S. A. Schumm, and D. E. Sugen: "Geomorphology," Methuen, New York, 1985.
Cloud, P.: "The Biosphere," *Sci. Amer.*, **249**(3), 176–189 (1983).
Cole, R. B., and A. R. Basu: "Middle Tertiary Volcanism During Ridge-Trench Interactions in Western California," 793 (October 30, 1992).
Courtillot, V. E.: "A Volcanic Eruption (and Mass Extinction)," *Sci. Amer.*, 85 (October 1990).
Craddock, C., et. al.: "Antarctic Geoscience," University of Wisconsin Press, Madison, Wisconsin, 1982.
Denton, G. H.: "The Last Great Ice Shects," Wiley-Interscience, New York, 1981.
Derry, D. R.: "A Concise World Atlas of Geology and Mineral Deposits," (Mining Journal Books, Ltd.), Halsted Press (John Wiley & Sons), New York, 1981.

Drake, C. L., and J. C. Maxwell: "Geodynamics—Where Are We and What Lies Ahead?" *Science*, **213**, 15–22 (1981).
Eather, R. H.: "Majestic Lights (Aurora)," American Geophysical Union, Washington, D. C., 1980.
Finkl, C. W., Jr.: "The Encyclopedia of Applied Geology," Van Nostrand Reinhold, New York, 1984.
Frankel, R. B., et. al.: "Magnetotactic Bacteria at the Geomagnetic Equator," *Science*, **212**, 1269–1270 (1981).
Gill, J., and M. Condomines: "Short-Lived Radioactivity and Magma Genesis," *Science*, 1368 (September 4, 1992).
Hallam, A.: "Great Geological Controversies," Oxford Univ. Press, New York, 1983.
Hanson, R. B.: "Planetary Fluids," *Science*, 281 (April 20, 1990).
Harland, W. B., et. al.: "A Geologic Time Scale," Cambridge Univ. Press, New York, 1983.
Haynes, C. V., Jr.: "Great Sand Sea and Selima Sand Sheet, Eastern Sahara: Geochronology of Desertification," *Science*, **217**, 629–633 (1982).
Head, J. W., and S. C. Solomon: "Tectonic Evolution of the Terrestrial Planets," *Science*, **213**, 62–75 (1981).
Hide, R., and J. O. Dickey: "Earth's Variable Rotation," *Science*, 629 (August 9, 1991).
Herbert, S.: "Darwin as a Geologist," *Sci. Am.*, **254**(5), 116–123 (1986).
Hess, H. H.: *Geol. Soc. Am. Buddington*, Vol. (1962), p. 599; J. T. Wilson, *Nature* (London) **207**, 343 (1965); and W. J. Morgan, *J. Geophys. Res.*, **73**, 1950 (1968).
Hoffman, K. A.: "Ancient Magnetic Reversals: Clues to the Geodynamo," *Sci. Amer.*, 76 (May 1988).
Hones, E. W., Jr., Ed.: "Magnetic Reconnection in Space and Laboratory Plasmas," American Geophysical Union, Washington, D. C., 1984.
Hones, E. W., Jr.: "The Earth's Magnetotail," *Sci. Amer.*, 40—47 (March 1986).
Hurlbert, S. H., and C. C. Y. Chang: "Ancient Ice Islands in Salt Lakes of the Central Andes," *Science*, **224**, 299–302 (1984).
Jeanloz, R.: "The Earth's Core," *Sci. Amer.*, **249**(3), 56–65 (1983).
Kerr, R. A.: "New Gravity Anomalies Mapped from Old Data," *Science*, **215**, 1220–1222 (1982).
Kerr, R. A.: "An Early Glacial Two-Step?" *Science*, **221,** 143–144 (1983).
Kerr, R. A.: "Continental Drilling Heading Deeper," *Science*, **224,** 1418–1420 (1984).
Kerr, R. A.: "First Look at the Deepest Hole," *Science*, **225,** 1461 (1984).
Kerr, R. A.: "Inyo Domes Drilling Hits Pay Dirt," *Science*, **227**, 504–506 (1985).
Kerr, R. A.: "Deep Holes Yielding Geoscience Surprises," *Science*, 468 (August 4, 1989).
Kerr, R. A.: "Having It Both Ways in the Mantle," *Science*, 1576 (December 4, 1992).
Lambeck, K.: "The Earth's Variable Rotation," Cambridge Univ. Press, New York, 1980.
Lapwood, E. R., and T. Usami: "Vibrations of the Earth," Cambridge Univ. Press, New York, 1981.
McCammon, C.: "Effect of Pressure on the Composition of the Lower Mantle End Member Fe_xO," *Science*, 66 (January 1, 1993).
McCauley, J. F., et. al.: "Subsurface Valleys and Geoarcheology of the Eastern Sahara Revealed by Shuttle Radar," *Science*, **218,** 1004–1020 (1982).
McKenzie, D. P.: "The Earth's Mantle," *Sci. Amer.*, **249**(3), 66–78 (1983).
Merrill, R. T., and P. L. McFadden: "Paleomagnetism and the Nature of the Geodynamo," *Science*, 345 (April 20, 1990).
Meyer, C.: "Ore Metals Through Geologic History," *Science*, **227**, 1421–1428 (1985).
Molnar, P.: "The Structure of Mountain Ranges," *Sci. Amer.*, **255**(1), 70–79 (1986).
Nance, A.: "Logging in the Red Zone," *Amer. Forests*, 6 (April 1982).
Newitt, L. R., and E. R. Niblett: "Relocation of the North Magnetic Dip Pole," Geological Survey of Canada Contribution No. 10786, 1062–1067, Earth Sciences, Energy, Mines and Resources Canada, Ottawa, Ontario, Canada (February 20, 1986).
Officer, C. B., and C. L. Drake: "The Cretaceous-Tertiary Transition," *Science*, **219**, 1383–1390 (1983).
Pardo, R.: "Rehabilitating St. Helens," *Amer. Forests*, 30 (November 1980).
Powell, C. S.: "Peering Inward" *Sci. Amer.*, 100 (June 1991).
Russell, C. T., and R. L. McPherron: "The Magnetotail and Substorms," *Space Science Reviews* **15,** 205–266 (December 1983).
Sakai, H., et al.: "Venting of Carbon Dioxide–Rich Fluid and Hydrate Formation in Mid-Okinawa Trough Backarc Basin," *Science*, 1093 (June 1, 1990).
Sedell, J. R., Franklin, J. F., and F. J. Swanson: "Out of the Ash," *Amer. Forests*, 20 (October 1980).
Siever, R.: "The Dynamic Earth," *Sci. Amer.*, **249**(3), 46–55 (1983).
Staff: "On Top of Kilauea, Geologists Predict Volcanic Activity," *Insight*, 2 (September 1987).
Staff: "Did Volcanism Overheat Ancient Earth?" *Science*, 746 (February 15, 1991).
Stebbins, J. F., and I. Farnan: "Nuclear Magnetic Resonance Spectroscopy in the Earth Sciences: Structure and Dynamics," *Science*, 257 (July 21, 1989).
Strangway, D. W., Ed.: "The Continental Crust and Its Mineral Deposits," Geological Association of Canada, Waterloo, Ontario, Canada, 1979.
Tarduno, J. A., et al.: "Rapid Formation of Ontong Java Plateau by Aptian Mantle Plume Volcanism," *Science*, 399 (October 15, 1991).
Valdridge, W. Scott, et al., Eds.: "Rio Grande Rift," New Mexico Geological Society, Socorro, New Mexico, 1984.
Vink, G. E., Morgan, W. J., and P. R. Vogt: "The Earth's Hot Spots," *Sci. Amer.*, 50 (April 1985).
Waldrop, M. M.: "Washington Embraces Global Earth Sciences," *Science*, **233**, 1040–1042 (1986).
Wells, S. G., and D. R. Haragan: "Origin and Evolution of Deserts," Colorado Mountain College, Glenwood Springs, Colorado, 1983.
Wohletz, K., and G. Heiken: "Volcanology and Geothermal Energy," Univ. of California Press, Berkeley, California, 1992.

Note: See also references listed at end of article on **Earth Tectonics and Earthquakes.**

EARTH AXIS. Any one of a set of mutually perpendicular reference axes established with the upright axis (the Z-axis) pointing to the center of the earth, used in describing the position or performance of an aircraft or other body in flight. The earth axes may remain fixed or may move with the aircraft or other object.

EARTH CURRENT. A large-scale surge of electric charge within the earth's crust, associated with a disturbance of the ionosphere. Current patterns of quasi-circular form and extending over areas the size of whole continents have been identified and are known to be closely related to solar-induced variations in the extreme upper atmosphere.

EARTH DAM. See **Dams.**

EARTHLIGHT. The illumination of the dark part of the moon's disk produced by sunlight reflected onto the moon from the earth's surface and atmosphere. Also called *earthshine.* Spectroscopic observations reveal that earthlight is relatively richer in blue light than is direct sunlight; this condition results from the fact that an appreciable part of the total earth reflection is backward-scattered light which, in accordance with the Rayleigh law, is relatively rich in the blue and poor in the red. See also **Moon (Earth's).**

EARTHMOVING EQUIPMENT. See **Coal.**

EARTH POINT. The point where the forward straight-line projection of a meteor trajectory intersects the surface of the earth.

EARTH RATE CORRECTION. A command rate applied to a gyroscope to compensate for the apparent precession of the gyro spin axis with respect to its base caused by the rotation of the earth.

EARTH-SHINE. See **Moon (The).**

EARTH TIDE. The response of the solid Earth to the forces that porduce the tides of the sea; semidaily earth tides have a fluctuation of beteeen 7 and 15 centimeters.

EARTH TECTONICS AND EARTHQUAKES. The Earth is an energetic, dynamic planet that is continuously, if ever so slowly, changing. Profound and extremely important changes have occurred since the planet was formed an estimated $4\frac{1}{2}$ billion years ago. Paleomagnetic and paleoseismic evidence, dating back only some one-half billion years, reveal how the shapes and juxtaposition of the continents and oceans have changed during the most recent 10% of the Earth's total life span to date.

Because these changes take place at such a *slow* rate, the human inhabitants of the Earth tend to develop the mistaken impression that, as they admire the magnificent landscape and seascape, the familiar features are forever, so to speak. The lifespan of a human is essentially *infinitesimal* on the time scale of the planet's changing events. Were it not for an occasional major earthquake or volcanic eruption, most persons would be oblivious of the gigantic forces and physicochemical processes at work under the Earth's crust.

The 1994 Northridge, California earthquake of magnitude 6.6+ (Richter scale) occurred in southern California in the San Fernando Valley, south of the San Gabriel Mountains and north of the city of Los Angeles. Thousands of aftershocks, some strong, occurred over a period of several days after the initial shock. Casualties were fewer than 100. Property damage was estimated in the billions of dollars. Great inconvenience, and interruption of normal living and the economy, of Greater Los Angeles resulted mainly from the creation of thousands of homeless citizens and extensive damage to the region's freeways. Although scientists did not regard the Northridge event as the "Big One," predicted for so many years, some seismologists turned their attention to the possibility of a number of previously undetected faults located directly under the downtown and outlying regions of Greater Los Angeles—sources of weakness unrelated to activity along the much-studied San Andreas Fault.

Studies of the Earth's early history fall within the broad province of *geochronology,* but are assisted in major ways by other applied sciences, such as seismology, radioactivity analysis, geomagnetics, probability and statistical analysis, volcanology, and orogeny, among many other technical disciplines.

Modern earthquake science and volcanology are largely guided by a relatively recent concept (1960s) known as *plate tectonics.* With this concept, scientists now have a basic physicochemical mechanism/process against which prior concepts can be tested and new hypotheses, working within the framework of plate tectonics, can be proposed. This is precisely what has been transpiring over the recent past—namely, that some aspects of the earlier assumptions of plate tectonics are being revised and refined with the help of new observational data. This research is being enhanced continuously through greatly improved field sensors, transmission systems, computer-assisted simulations, and laboratory techniques for determining the ultimate physical properties, such as the use of the diamond anvil high pressure cell, and also by the high temperatures achieved with laser beams.

Plate Tectonics

Although the theory of plate tectonics is a comparatively recent development, it followed a slow but logical development, in essence, since the early mapmakers of the Earth, dating back to the 1400s and 1500s (Period of the Great Explorers).

Concept of Continental Drift. Obviously, early mapmakers were persons of keen curiosity and advanced intellect for their time. As maps of the Earth were developed, particularly those involving most of the oceans and continents, questions arose as to why the boundaries of the continents and oceans essentially had matching images (i.e., the eastern borders of a continent tended to match (in jigsaw puzzle fashion) the western borders of another continent). This would have been particularly noticeable in the matching of the eastern edge of South America with the western edge of Africa. Such matching could not be ascribed simply to coincidence or chance. However, years went by and it was not until Francis Bacon, in 1620, made formal mention of the provocative geographical correspondences, but he did not theorize on a possible cause. A few other curious scholars, however, introduced some very generalized scenarios, including the case of sinking *Atlantis.*[1] Much prior to the full development of the plate tectonic theory, a visionary meteorologist, Alfred Wegener of Berlin, in 1912 proposed the hypothesis of *continental drift.* At that time, Wegener's concept was not amenable to the then prevailing theory of the structure and maturing of the Earth. The model[2] then popular claimed that, after the Earth was formed from a molten state, it cooled, becoming solid inside—with heavy elements, such as iron, sinking to form a solid core and lighter elements rising to form a surface crust upon cooling. Essentially accurate regarding some aspects, the model did not explain the mechanism of continent and ocean formation, nor how the joining of or the drifting apart of continents affected the ultimate tectonic nature of the Earth's surface. Wegener did find considerable evidence for his hypothesis in the form of "matching" fossils, animals, and geological features. See Fig. 1.

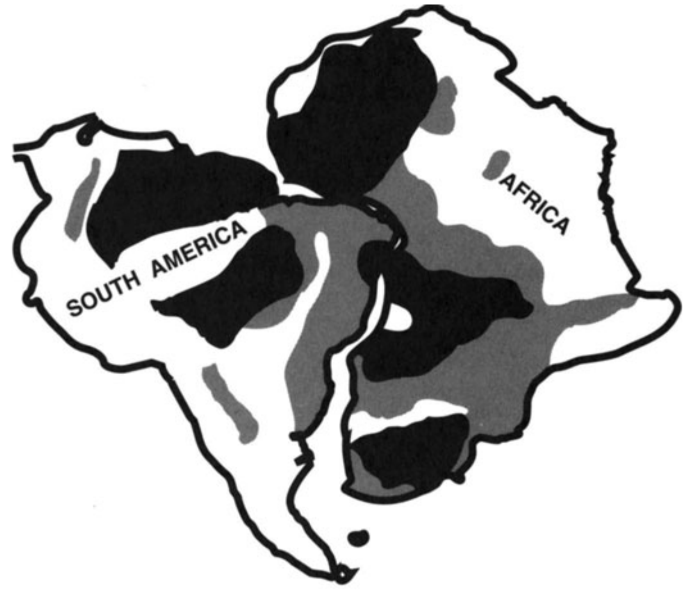


Fig. 1. Very approximate depiction of how Wegener envisioned Africa and South America were nestled prior to slowly drifting apart. Black areas represent regions of continental structure that ultimately broke during the separation of the continents. Similarly, regions of crust (shown in gray), most of which also indicate as being part of essentially a continuous land mass prior to drifting away from each other. Wegener demonstrated quite well his premise of continental drift, but did not explain the mechanism behind the movement of continents. Thus, his theory did not embrace the full concept of plate tectonics. [*Diagram is patterned after A. Hallam. (See reference)*].

Some contemporary researchers feel that Wegener came pretty close to developing the plate theory, but was limited by a lack of precise instruments, inadequate surveying techniques, and to a degree, he was just a few decades ahead of his peers. In the excellent A. Hallam article (see reference), there is a review of Wegener's lifetime contribution to geophysical science.

Distribution of the continents over many millenia of time has been suggested by a number of researchers. One concept suggested by Wegener was that, at one time in the Earth's development, land masses comprised a supercontinent, which he referred to as *Pangea.* See Fig. 2.

Plate Tectonics Simplified. The Earth model based on the concept of plate tectonics describes the surface of the planet (biosphere, ecosphere, etc.) as supported by underlying large, broad, and thick plates that are composed of continental and oceanic crust and materials from the upper mantle. Because, in essence, these plates "float" on an underlying viscous material of the mantle, they have the ability to move, much as a ship may move on the ocean. Thus, unlike a soundly engineered structure, such as a building or bridge, which is supported by a strong, reliable foundation upon which the structural integrity of the building depends, the foundation supporting the Earth's surface layer is not so rigid and dependable. To carry the analogy one further step, one

[1] A fabled island in the Atlantic that, according to legend, was submerged beneath the sea.

[2] Developed by G. K. Gilbert (United States Geological Survey) and H. Reed (Johns Hopkins Univ.), 1906.

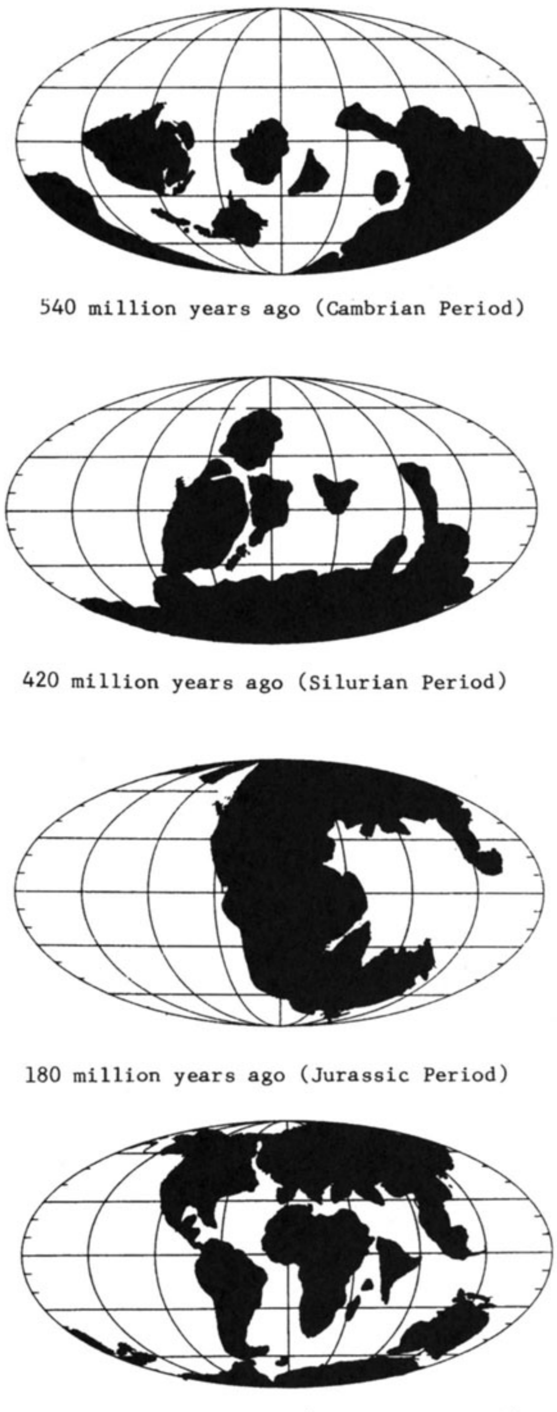


Fig. 2. Abridged and simplified facsimiles of the distribution of continents and oceans on Earth based upon paleomagnetic evidence dating back 540 million years. A series of more detailed, computer-generated maps was prepared by researchers in connection with the University of Chicago Paleographic Atlas Project. (*See also Siever (1983) reference*.)

may visualize a large building, one side of which rests on a movable platform and the other side of the building which rests on another movable platform. The platforms may move in the same general direction at different rates; or they move in opposing directions. In any case, if the platforms do not move in complete synchronism (a condition not found in plate tectonics), the building will either move in on itself, devastatingly crushing one portion against the next; or parts of the building will move away from each other, causing equal damage.

Continents in total or in part are supported by separate plates as indicated by the detailed map of Fig. 3. The plates move more or less independently and at rates measured in terms of a few centimeters per year. Like ships, the plates tend to bump one another, one plate rising over or subducting below an adjacent plate. They may grind away at each other like ice floes in a river. They are driven by energy coming from the phenomenon know as *seafloor spreading*, which is the hypothesis that the oceanic crust is increasing by convective, upwelling of magma along the mid-oceanic ridges or world rift system and moving away the newly created material at a rate of from 1 to 10 centimeters per year. See **Oceanography.**

The forces moving the plates, although acting slowly, are relentless and where plates push together or pull apart, the strains set up in the materials so contacted build up steadily and slowly to very high values. Periodically, the time is reached when so much energy no longer can be stored (as strain) in the materials and fractures thus occur. Sudden bursts of released energy cause earthquakes. The manner in which plates interact to form volcanoes is similar and is described in the article on **Volcano.**

As indicated by Fig. 3, most of the Earth's seismic activity occurs at the boundaries of plates. The greatest seismic activity, the largest shocks, and the deepest shocks, occur at places where plates converge (the arcs such as Japan and Tonga), where one plate is thrust beneath another to depths at least as great as the depths of the deepest earthquakes. Where plates diverge (as along the Mid-Atlantic Ridge), or slide past one another (as along the San Andreas Fault in California), seismic activity is shallow, and although substantial, is usually not as great as that of the arcs. The global patterns of the focal mechanisms of earthquakes also fit, in general, the patterns of plate motion. Seismological evidence played an important role in the development of the concept of plate tectonics. See Fig. 4.

It has been observed that most strong-motion earthquake data have been accumulated from events in California and Japan. California and Japan represent two different tectonic regions. The west coast of the United States is not typical of continental margins. The primary seismic feature is the well-known San Andreas fault, with its right lateral strike-slip movement (known as a transform fault), which is usually associated with mid-oceanic spreading ridges where it is of much shorter length and a consequence of the actual three-dimensional nature of the plates moving on the surface of the earth. The San Andreas fault joins the spreading East Pacific Rise and the spreading zones of the Gorda and Juan de Fuca ridges to the north. The realization that the fault was of a transform type was first made by J. Tuzo Wilson (University of Toronto). Wilson recognized that a fault displacement of up to 1000 kilometers (620 miles) required some mechanism that allowed the displacements to occur while satisfying the laws for the conservation of matter. Because of the special nature of the relative movements between the Pacific Plate and the North American Plate, the San Andreas fault may be considered as an extension and a very large example of a mid-oceanic feature.

The majority of continental margin seismic activity stems from direct plate collisions, as explified by the activity along the Chilean coast. The focal depths, which are shallow off the coast, become progressively deeper as activity occurs further inland on the eastern side of the Peru-Chile trench.

The mechanism of seafloor spreading, considered to be the source of energy for moving the tectonic plates, is under continuous scrutiny by numerous geoscientists. Oceanographic investigations are revealing much new information on spreading and the formation of ocean ridges and trenches. Such research is exemplified by the current understanding of the nature and mechanics of the Cayman's Zigzag Rift. See Fig. 5. The actual mechanics of crustal formation are poorly understood. Early acoustic sound methods revealed very large zones of deep fracturing, fractures that stretch almost the width of an ocean. These form perpendicular to a ridge wherever one section of the ridge crest is offset horizontally from another. Fracturing also occurs on a very small scale. Such fracturing has been revealed by new instruments, such as side-scanning sonar. Small-scale cracking seems to result from thermal stresses during cooling. Johnson (University of Texas at Dallas) counted the cracks in a DSDP core recovered from 110-million-year-old crust in the Atlantic near Bermuda. It was found that the typical crack in the core was about 2 millimeters wide and 150 millimeters long, and that such cracks occurred about every 2 centimeters. Zones of intense cracking were found to occur through a 247-meter (810 foot) core. The end result of all this fracturing appears to be that the crust on and near the ridge crest is very permeable to water, possibly on the order of 20–25%.

Earthquakes

An earthquake may be defined as a sudden motion or trembling in the earth caused by the abrupt release of slowly accumulated strain. In certain locations of the world, earthquakes are to be expected—they are the rule, not the exception, as natural consequences of plate tectonics. Numerous variables affect the timing, the extent (magnitude) and exact location of an earthquake—hence they are difficult to predict. However, progress is being made. Synonyms for earthquake include: *Temblor* (sometimes spelled tremblor), *seism, macroseism* (as opposed to microseism), and *shock*.

Characteristics of Earthquakes

Aftershock. An earthquake which follows a larger earthquake or main shock and originates at or near the focus of the larger earthquake. Generally, major earthquakes are followed by a larger number of aftershocks, decreasing in frequency with increasing time. A series of aftershocks may last many days after small earthquakes; even months after large earthquakes.

Amplitude. This varies from a fraction of a centimeter to several centimeters. The destructive phase of an earthquake varies from one minute to but a few seconds. It is estimated that over the Earth there are over a million earthquakes each year, the majority of them occurring in regions of recent mountain building. The majority are quite weak. It is estimated that a major earthquake occurs on the earth about once per week. Tremors of the ocean bottom cause seismic seawaves (tsunamis). These seawaves have been known to rise 30 meters (100 feet) or more. When such waves break upon a densely inhabited coast, there is great loss of life and destruction of property. A list of major earthquakes in terms of lives lost is given in Table 7 under **Earth.**

Excluding certain near-surface regions, the velocities of dilational seismic waves range from about 5 kilometers per second in parts of the crust to a maximum of 13.5 kilometers per second at the base of the mantle; corresponding shear wave velocities range from 3 to 8 kilometers per second. The shortest periods of interest in the study of waves from distant earthquakes are of the order of $\frac{1}{3}$ second frequency (frequency = 3 Hz); the longest periods are about 53 minutes (frequency = ~ 1 cycle per hour), and they correspond to a free oscillation of the earth in the fundamental spheroidal mode.

Free oscillations of measurable amplitudes are generated only by the largest earthquakes. The largest earthquakes probably release between 10^{24} and 10^{25} ergs in the form of seismic waves, and the few largest shocks account for most of the energy released in this form. Mean annual release is estimated at 9×10^{24} ergs.

Deep Earthquake. In most earthquakes, the Earth's crust cracks like porcelain. Stress builds up until a fracture forms at a depth of a few kilometers and *slip* rock structure relieves the stress. A deep earthquake occurs hundreds of kilometers below the Earth's surface, i.e., down into the mantle where high pressure is believed to prevent rocks from cracking. Since 1989, more than 60,000 earthquakes at depths greater than 70 km (22% of all earthquakes) have been recorded. The subduction zone is the setting for nearly all deep earthquakes. There are numerous hypotheses pertaining to where and how the shock is created. One concept is based upon a sudden change of state, as temperature and pressure change with depth, from a plastic medium (loaded with energy) to a crystalline material capable of fracturing. Research in this area is in a comparatively early stage, and advantage is being taken of observing materials change under great pressures, as accomplished by the diamond anvil, and at very high temperatures, as effected by laser beams.

Earthquake Swarm. A series of minor earthquakes, none of which may be identified as the main shock, occurring in a limited area and time, frequently in the vicinity of a volcano.

Epicenter. That point on the Earth's surface directly above the focus of an earthquake.

Fault. A surface or zone of rock fracture along which there has been displacement from a few centimeters to a few kilometers in scale.

Foreshock. A tremor that commonly precedes a larger earthquake or main shock by seconds to weeks (possibly years) and that originates at or near the focus of the larger earthquake.

Focal Sphere. An arbitrary reference sphere drawn about the focus of an earthquake, to which body waves recorded at the Earth's surface are projected for studies of earthquake mechanisms.

Focus. That point within the Earth which is the center of an earthquake and the origin of its elastic waves. Sometimes called *hypocenter*.

Hidden Earthquake. A seismic event that occurs on a blind fault under folded terrain.

Isostasy. The state of equilibrium within the Earth's crust that is buoyantly supported by the plastic materials in the mantle.

Magnitude/Energy. See later paragraphs in this article.

Seismic. A term used to describe anything pertaining to an earthquake or earth vibration.

Seismograph. An instrument for recording vibrations of the Earth, particularly of earthquakes or artificially induced energy for the exploration of underlying rock formations and the interior of the Earth. The record produced by a seismograph is a *seismogram*.

Seismography. The study of the theory of seismographs.

Seismology. The study of earthquakes and, by extension, the investigation of the depths of the Earth (as in the case of oil and gas exploration) by way of measuring and analyzing natural or artificially generated seismic signals.

Strike-slip. Movement parallel with the strike of a fault, Subduction Zone.

Strike-slip Fault. A fault showing predominantly horizontal movement parallel to the strike; an absence of vertical displacement.

Subduction Zone. Area of crustal plate collision where one crustal block descends beneath another, marked by a deep ocean trench caused by the bend in the submerging plate. Downward movement of the subducting plate results in earthquakes, volcanoes, and intrusions on the far side of the trench.

Tremor. A minor earthquake, specially a foreshock or an aftershock.

Tsunami. A gravitational seawave produced by any large-scale, short-duration disturbance of the ocean floor, principally by a shallow submarine earthquake, but also by submarine earth movement, subsidence, or volcanic eruption, characterized by great speed of propation (up to 950 km/hr), long wavelength (up to 200 km), long period (varying from 5 minutes to a few hours), generally 10 to 60 minutes, and low, observable amplitude on the open sea, although it may pile up to great heights (30 meters or more) and cause considerable damage on entering shallow water along an exposed coast thousands of kilometers from the source. Also known as a seismic sea wave.

Faults. Generally considered to be a great fracture of the earth along which movement has taken place, with the result that the crustal blocks are displaced relative to one another. This movement, in most cases, appears to be intermittent and the actual individual displacements may be very small, but by accumulation may reach tens, hundreds, or rarely thousands of feet (meters). The fractures themselves may often be traced for many miles. The San Andreas fault, horizontal movement along a portion of which caused the earthquake that so severely damaged San Francisco, California, in 1906, has been traced for about 600 miles (965 kilometers).

In describing faults (See Fig. 6), certain terms have been adopted. Those in most common use are given herewith: the *fault plane* is the plane of the fracture and may be vertical or at some other angle; the angle between the fault plane and the horizontal is called the *angle of dip of the fault plane*. The angle between the vertical and the fault plane is spoken of as the angle of hade or simply as the hade of the fault. The surfaces of the fault plane are called the *walls of the fault*. If the fault plane dips, the uppermost wall is called the *hanging wall*, the lower wall the *foot wall*. These terms are applied irrespective of whether the fault is normal, with the hanging wall slipping down the dip of the fault plane, or whether it is a reverse fault with the hanging wall apparently pushed up the dip of the fault plane. A normal fault is sometimes spoken of as a *gravity fault*. The displacement measured along the fault plane is designated as the *slip*; the displacement measured vertically is called the *throw*; the displacement measured at right angles to the plane of the involved stratum is called the *stratigraphic throw*. The amount of horizontal displacement between the ends of a broken stratum measured at right angles to the direction of strike of

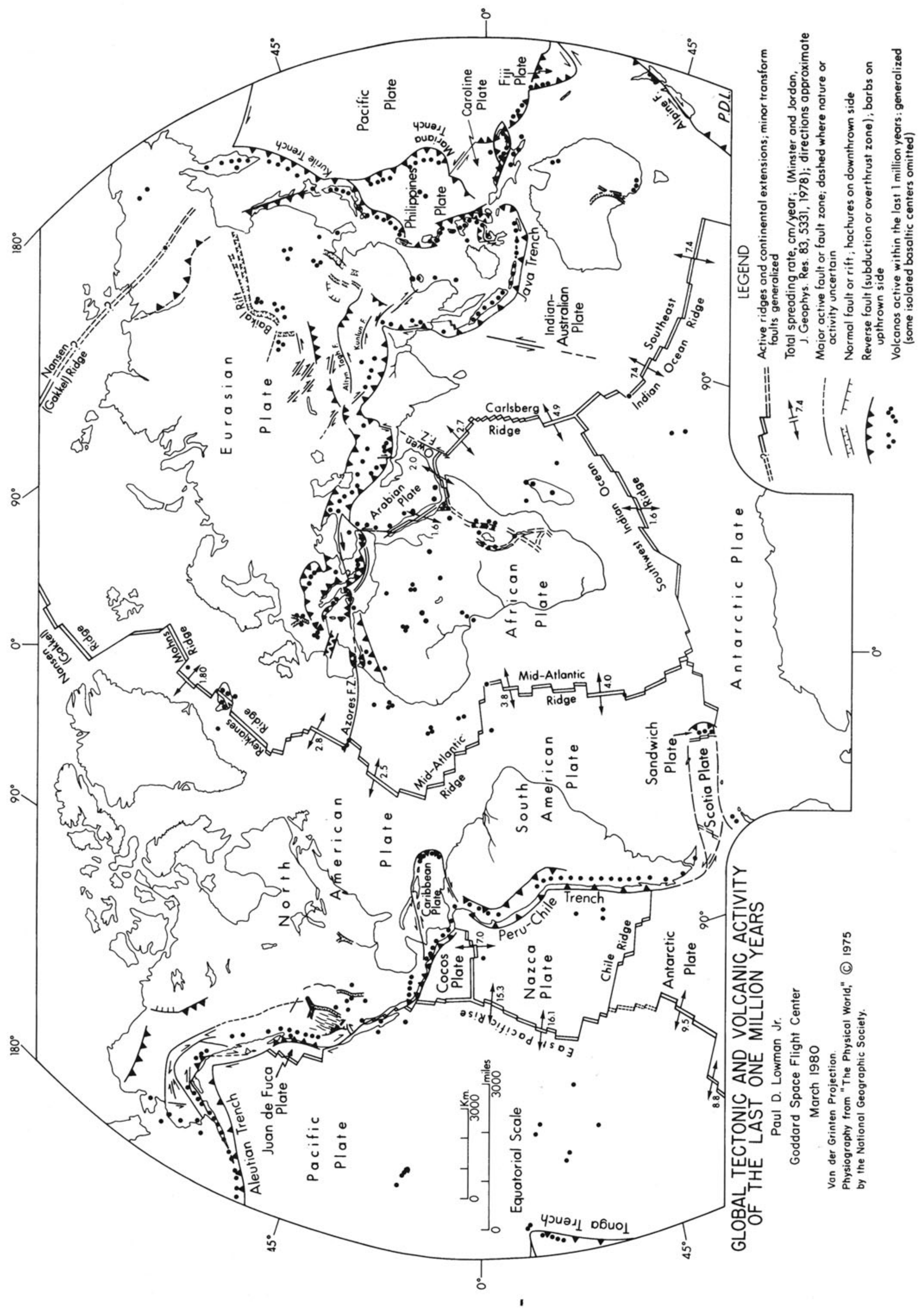

GLOBAL TECTONIC AND VOLCANIC ACTIVITY OF THE LAST ONE MILLION YEARS
Paul D. Lowman Jr.
Goddard Space Flight Center
March 1980
Van der Grinten Projection.
Physiography from "The Physical World," © 1975 by the National Geographic Society.
LEGEND
Active ridges and continental extensions; minor transform faults generalized
Total spreading rate, cm/year; (Minster and Jordan, J. Geophys. Res. 83, 5331, 1978); directions approximate
Major active fault or fault zone; dashed where nature or activity uncertain
Normal fault or rift; hachures on downthrown side
Reverse fault (subduction or overthrust zone); barbs on upthrown side
Volcanos active within the last 1 million years; generalized (some isolated basaltic centers omitted)
Equatorial Scale
Km. 0 3000
miles 0 3000
Eurasian Plate
Pacific Plate
North American Plate
South American Plate
African Plate
Antarctic Plate
Indian-Australian Plate
Arabian Plate
Philippines Plate
Caroline Plate
Fiji Plate
Juan de Fuca Plate
Cocos Plate
Nazca Plate
Caribbean Plate
Sandwich Plate
Scotia Plate
Aleutian Trench
Kurile Trench
Mariana Trench
Java Trench
Peru-Chile Trench
Tonga Trench
Nansen (Gakkel) Ridge
Mohns Ridge
Reykjanes Ridge
Mid-Atlantic Ridge
Azores F.Z.
Carlsberg Ridge
Owen F.Z.
Southeast Indian Ocean Ridge
Southwest Indian Ocean Ridge
East Pacific Rise
Chile Ridge
Baikal Rift
Altyn Tagh F.
Kunlun F.
Alpine F.
P.D.L.
0°
45°
90°
180°

Fig. 3. Principal plates, ridges, and trenches according to the theory of global plate tectonics. The outer surface of the earth, just under the crust and known as the lithosphere, supports a number of rigid plates. These plates may be about 100 kilometers (62 miles) or more in thickness. The continents (about 40 kilometers; 24 miles thick) essentially rest on the larger lithospheric plates. In the case of the western United States, southern California is part of the Pacific Plate, whereas the rest of North America is on the North American Plate. The three main forms of plate boundaries are (1) mid-oceanic ridges, (2) transform faults, and (3) subduction zones. New crust is created along the mid-oceanic ridges by the rise of basaltic magma. This cools and is carried away by sea-floor spreading. The mechanism of seafloor spreading was conceived as the result of two important discoveries: (1) the mid-oceanic ridge system, and (2) the zebra-stripe pattern of magnetic anomalies which symmetrically flank these ridges. Studies of ocean floor rocks support a presumption that the sea floors are moving away from the mid-oceanic ridges. Transform faults, along which the mid-oceanic ridges are offset, provide areas where the lithospheric plates can slide past one another. Subduction zones are the sites of destruction of the lithospheric plates. There are places in the ocean trenches where one oceanic plate descends beneath another. Should a second plate have a continental crust, then mountains would be inclined to develop along the continental margin, as exemplified by the Andes. Where two continental crustal plates collide, even greater mountain-building activity is created. Movement of plates so affects the crust that earthquakes, mountain building, and volcano formation become an intimate part of the system. A majority of volcanoes occur in arcs which are parallel to the oceanic trenches and are believed to result from one ocean crustal plate being subducted beneath another plate. Partial melting of the oceanic crust on the descending plate may occur as it passes onto the high-temperature athenosphere. Thus, magma ascends to the surface and forms volcanoes and underlying granitic batholiths. Volcanoes are also situated along the mid-oceanic ridges and, less frequently, over isolated "hot spots" in the athenosphere, exemplified by the Hawaiian volcanoes. See **Volcano.**

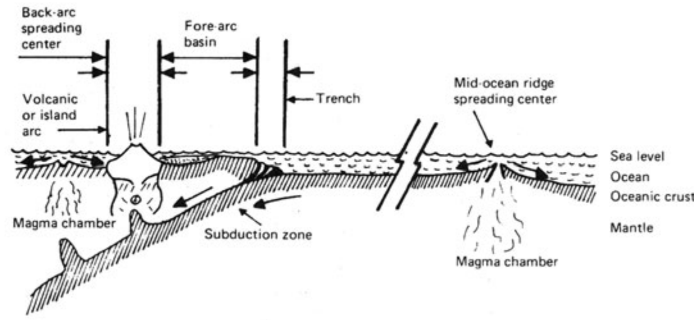


Fig. 4. Important features of plate tectonics. Examples are taken from the Mariana island-arc system. At depths of 90–100 kilometers (56–62 miles), the descending slab commences to melt and magma rises toward the surface, forming volcanic island arcs. The depressed area between the volcanic arc and trench gradually fills with sediments forming the *fore-arc basin*. Other magmas rise to the surface along the Mariana *back-arc basins*, creating a *spreading center* along which new crustal material forms at the surface. *Island arc*: Also known as a volcanic arc, an island arc is usually a chain of islands, e.g., the Aleutians, rising from the deep sea floor and near to the continents. *Trench*: A narrow, elongated depression of the deep-sea floor, with steep sides and oriented parallel to the trend of the continent and between the continental margin and the abyssal hills. Such a trench may be 2 or more kilometers (1.2 miles) or deeper than the surrounding ocean floor, and may be thousands of kilometers long. See also **Earth;** and **Ocean.** (*Sketch suggested by Davin and Gross, 1980.*)

the fault plane, is called the *heave*. The visible evidence of the trace of a fault plane at the earth's surface is called the *fault trace*. The block of the earth's crust which has moved downward, relatively speaking, to the other is called the *downthrown block* or referred to as the *downthrow side* of the fault. The other block is called the *upthrown block* or the *upthrow side* of the fault. If the strike of the fault plane is essentially at right angles to that of the bedding it is called a *dip fault*. A *strike fault* is one in which the movement has been parallel to the strike of the strata involved. A *compound fault* involving several parallel displacements dipping in the same direction, resulting in a step-like arrangement, is referred to as a *step fault*. The term *graben* or *trough fault* refers to a downthrown area bounded on each side by two or perhaps more faults. A *horst* is an uplifted area bounded by two or more faults. See Fig. 7.

Considerably more information on the characteristics of faults in various regions of Earth is given later in this article.

Public Reactions to Earthquakes

The majority of earthquakes that occur on the Earth are not detectable by people. In relating Richter Scale magnitudes to the size and energy of an earthquake, the following figures apply:

4.5	Detectable within 20 miles of epicenter. Possible slight damage within a small area.
6	Moderately destructive.
7	A major earthquake.
8	A great earthquake.

The frequency of occurrence of earthquakes increases by about a factor of 8 or 10 per unit of magnitude as *the magnitude decreases*. On the average, only about 25 shocks with a magnitude of 7 or more occur each year, but it has been estimated that there are at least one million earthquakes per year, most of them quite small. About 5,000 to 10,000 of these are routinely located and studied.

Quite understandably, earthquakes that cause major losses of life and property provoke the greatest attention. However, earthquakes of even greater magnitude may occur in relatively isolated areas which are principally of concern to seismologists.

Three parameters are important in assessing an earthquake: (1) Duration of the earthquake; (2) velocity of the surface movement; and (3) the rate of change of this velocity. In their potential for damage, these three factors are closely related. Some of these relationships are immediately obvious. A very short earthquake of high velocity—only one or two cycles of ground motion—is less damaging than an earthquake causing similar motion for many cycles. An earthquake with high ac-

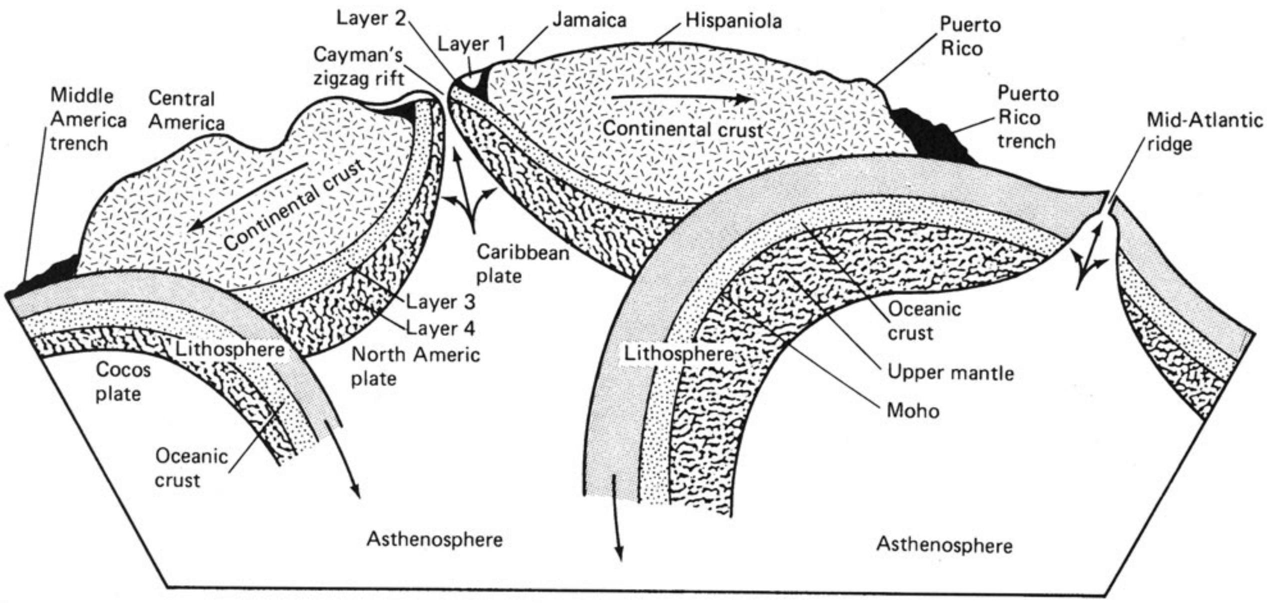


Fig. 5. The Cayman Trough off eastern Central America represents a great tear or shear in the ocean's crust. The trough is some 1500 kilometers (3052 miles) long and about 4 times the depth of the Grand Canyon. It is growing as two plates of the earth's crust (North American Plate and Caribbean Plate) slide past each other, but at a very slow pace. This tectonic action has created very large seismic events, including the February 1976 Guatemalan earthquake that killed some 23,000 people. As the plates separate, magma from the interior slowly rises into the rift valleys, thus continuously building oceanic crust. When plates collide, one edge plunges beneath the other, to be reabsorbed in the interior. Often volcanoes will be found along the overriding edge. Not to scale.

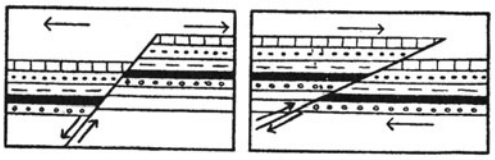


Fig. 6. Structure sections of a single normal fault (left); and simple thrust fault (right).

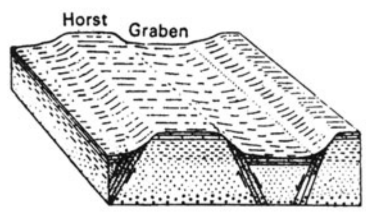


Fig. 7. Block diagram of one type of scenery produced by "normal" or block faulting. It may also be assumed that the graben has been pushed down, and the horst has been pushed up. Neither structure is necessarily entirely the result of tension. Note that the stratigraphic (vertical) order of the formations has not been duplicated or reversed.

celeration but low velocity is less damaging than one causing higher velocities. As part of an effort to develop seismic design standards for buildings and structures, these factors have been combined in a map which tentatively characterizes potential earthquake risk in terms of acceleration and velocity throughout the United States. Participants in this development were the American Society of Civil Engineers, the University of California (Seismographic Station, Berkeley), and the Massachusetts Institute of Technology (Department of Civil Engineering). Efforts are sponsored by the National Science Foundation and the U.S. National Bureau of Standards. See Fig. 8.

As early as 1883, the Italian geologist, M. S. De Rossi, and the Swiss naturalist, F. A. Forel developed a scale of one to ten for expressing the intensity of an earthquake. Earthquake intensity may be defined as a measure of the *effects* of an earthquake, notably the effects in terms of people and structures. Earthquake intensity not only will be dependent upon the strength (or magnitude) of the earthquake, but also upon the distance from the epicenter. Intensity also will be markedly affected by local geology, by the numbers and kinds of structures in a given area, as well as the concentration of people within the affected area. Even the time of day may have a large bearing upon the effects, with large numbers of people assembled in factories, schools, offices, etc., during daytime hours. From a scientific standpoint, the *effects* may be of less value than measurements of *magnitude*. However, from an engineering and technological viewpoint, great knowledge has been learned from studying the aftermath of earthquakes, particularly in terms of structures of all types.

The De Rossi-Forel one-to-ten scale was ultimately replaced by the Mercalli scale, devised by Giuseppi Mercalli, an Italian geologist, in 1902. Again, this was an arbitrary scale of earthquake intensity, ranging from I (detectable only instrumentally) to XII (causing almost complete destruction). This scale was later adapted to North American conditions and became known as the *modified Mercalli scale*. This modified scale (MM) was prepared in 1931 by American seismologists, H. O. Wood and F. Neumann and takes into consideration such features as tall buildings, motor cars and trucks, and underground pipes, not included in the earlier, unmodified scale.

Although not purely based on scientific data, the modified Mercalli Scale is an excellent summary, not only of the human reactions to an earthquake but also of important structural events that usually take place with the increasing magnitude of an earthquake. It is also a good guide for architects and planners when designing new structures and modifying old structures, as well as for organizations that provide emergency services. See Table 1.

Seismology

Many seismographs are of the inertial type, depending upon measurement of the relative displacement between a point fixed to the Earth and a mass loosely coupled to the Earth. Other instruments measure relative displacement between two points on the Earth. Seismographs installed throughout the world now number in the several thousands. Seismic instrumentation also includes accelerometers. See article on **Accelerometer.**

TABLE 1. AN ORGANIZED SUMMARY OF HOW PEOPLE AND STRUCTURES RESPOND TO EARTHQUAKES OF INCREASING MAGNITUDE
(Based on the Modified Mercalli Scale and given in order of increasing intensity)

1. *Not felt by persons, except under particularly favorable circumstances, Dizziness or nausea, however, may be experienced.*
Birds and animals may appear uneasy, disturbed. Trees, structures, liquids, bodies of water may sway gently: doors may swing slowly.

2. *Detected indoors by a few persons, particularly on upper floors of a multistory building; and by sensitive or nervous persons.*
Same manifestations as I, plus hanging objects will swing if delicately suspended.

3. *Detected indoors by several persons, usually as a rapid vibration, but which may not be recognized as an earthquake immediately. Vibration is similar to that from a passing, lightly-loaded truck, or of a heavily-loaded truck from some distance. In some instances, duration is sufficient to be estimated.*
Manifestation of II, plus appreciable movements on upper levels of tall structures. A standing motor car may rock slightly.

4. *Detected indoors by many persons and outdoors by a few persons. Awakens a few individuals, particularly light sleepers, but experience is not frightening except by persons who have had prior experience with earthquakes. Vibration similar to that from the passing of a heavily loaded truck. Sensation is like that of a heavy body striking the structure, or the falling of heavy objects inside the structure.*
Dishes, windows, and doors rattle: glassware and crockery clink and may crash. Walls and structure frame (particularly of a wooden house) creak, especially if intensity is in the upper range of this class. Hanging objects swing. Liquids in open vessels are disturbed. Parked automobiles rock noticeably.

5. *Detected indoors by practically everyone: and outdoors by most everyone. Slight excitement; some persons may run outdoors. Awakens most sleepers. Frightens a few persons.*
Buildings tremble throughout. Dishes and glassware break to some extent. Windows crack in some cases, but not generally. Vases and small or unstable objects frequently overturn. Hanging objects and doors swing generally or considerably. Pictures knock against walls or swing out of position. Doors and shutters open or close abruptly. Pendulum clocks stop or run fast or slow. Small objects move, and furnishings may shift to a slight extent. Small amounts of liquids spill from well-filled open containers. Trees and bushes shake slightly.

6. *Detected by everyone, indoors and outdoors. Awakens all sleepers. Frightens many people: general excitement, and some persons run outdoors.*
Persons move unsteadily. Trees and bushes shake slightly to moderately. Liquids are set in strong motion. Small bells in churches and schools ring. Poorly built buildings may be damaged. Plaster falls in small amounts. Other plaster cracks somewhat. Many dishes and glasses break, and a few windows break. Knick-knacks, books and pictures fall. Furniture overturns in many instances. Heavy furnishings move.

7. *Frightens everyone. General alarm, and everyone runs outdoors.*
People find it difficult to stand. Persons driving cars notice shaking. Trees and bushes shake moderately to strongly. Waves form on ponds, lakes, and streams. Water is muddied. Gravel or sand stream banks cave in. Large church bells ring. Suspended objects quiver. Damage is negligible in buildings of good design and construction; slight to moderate in well-built ordinary buildings; considerable in poorly built or badly designed buildings, adobe houses, old walls (especially where laid up without mortar), spires, etc. Plaster and some stucco fall. Many windows and some furniture break. Loosened brickwork and tiles shake down. Weak chimneys break at the roofline. Cornices fall from towers and high buildings. Bricks and stones are dislodged. Heavy furniture overturns. Concrete irrigation ditches are considerably damaged.

8. *General fright and alarm approaches panic.*
Persons driving cars are disturbed. Trees shake strongly, and branches and trunks break off (especially palm trees). Sand and mud erupts in small amounts. Flow of springs and wells is temporarily and sometimes permanently changed. Dry wells may renew flow. Damage slight in brick structures built especially to withstand earthquakes; considerable in ordinary substantial buildings, with some partial collapse; heavy in some wooden houses with some tumbling down. Panel walls break away in frame structures. Decayed pilings break off. Walls fall. Solid stone walls crack and break seriously. Wet ground and steep slopes crack to some extent. Chimneys, columns, monuments, and factory stacks and towers twist and fall. Very heavy furniture moves conspicuously or overturns.

9. *Panic is general.*
Ground cracks conspicuously. Damage is considerable in masonry structures built especially to withstand earthquakes; great in other masonry buildings—some collapse in large part. Some wood frame houses built especially to withstand earthquakes are thrown out of plumb; others are shifted wholly off foundations. Reservoirs are seriously damaged, and underground pipes sometimes break.

10. *Panic is general.*
Ground, especially when loose and wet, cracks up to widths of several inches; fissures up to a yard in width run parallel to canal and stream banks. Landsliding is considerable from river banks and steep coasts. Sand and mud shifts horizontally on beaches and flat land. Water level changes in wells. Water is thrown on banks of canals, lakes, rivers, etc. Dams, dikes, embankments are seriously damaged. Well-built wooden structures and bridges are severely damaged, and some collapse. Dangerous cracks develop in excellent brick walls. Most masonry and frame structures, and their foundations, are destroyed. Railroad rails bend slightly. Pipelines buried in earth tear apart or are crushed endwise. Open cracks and broad wavy folds open in cement pavements and asphalt road surfaces.

11. *Panic is general.*
Disturbances in ground are many and widespread, varying with the ground material. Broad fissures, earth slumps, and land slips develop in soft, wet ground. Water charged with sand and mud is ejected in large amounts. Sea waves of significant magnitude may develop. Damage is severe to wood frame structures, especially near shock centers; great to dams, dikes and embankments, even at long distances. Few if any masonry structures remain standing. Supporting piers or pillars of large, wellbuilt bridges are wrecked. Wooden bridges that "give" are less affected. Railroad rails bend greatly, and some thrust endwise. Pipelines buried in earth are put completely out of service.

12. *Panic is general.*
Damage is total, and practically all works of construction are damaged greatly or destroyed. Disturbances in the ground are great and varied, and numerous shearing cracks develop. Landslides, rock falls, and slumps in river banks are numerous and extensive. Large rock masses are wrenched loose and torn off. Fault slips develop in firm rock, and horizontal and vertical offset displacements are noticeable. Water channels, both surface and underground, are disturbed and modified greatly. Lakes are dammed, new waterfalls are produced, rivers are deflected, etc. Surface waves are seen on ground surfaces. Lines of sight and level are distorted. Objects are thrown upward into the air.

During the past few decades, there has been a marked increase toward improving seismic measurements, including the research instrumentation used when studying the ocean floor and submerged mountains and valleys. Although it was about a quarter of a century ago, it is interesting to note that, during the Apollo Program, seismographs were placed on the moon for study of that satellite.

There are governmental and private organizations which exchange seismographic information on a worldwide basis. These include the World Wide Seismic Network (WWSN), established in the early 1960s, the Lamont-Doherty Geological Observatory at Columbia University, the Coordinating Committee for Earthquake Prediction (CCEP), the U.S. Geological Survey (Menlo Park, California), the California Institute of Technology, the American Institute of Civil Engineers, the Woods Hole Oceanographic Institution (Woods Hole, Massachusetts), the Scripps Institution of Oceanography (La Jolla, California), and the Japan Meteorological Agency, among others.

The operating principle of a seismograph is shown in Fig. 9. When the ground shakes, the suspended weight, because of its inertia, scarcely moves, but the shaking motion is transmitted to the marker, which leaves a record on the drum. There are, of course, several configurations which utilize this basic operating principle. A representative seismogram is shown in Fig. 10. This is the record of an earthquake of magnitude 6.5 (Aleutian Islands). The record was made at a station some 4,000 kilometers from the epicenter.

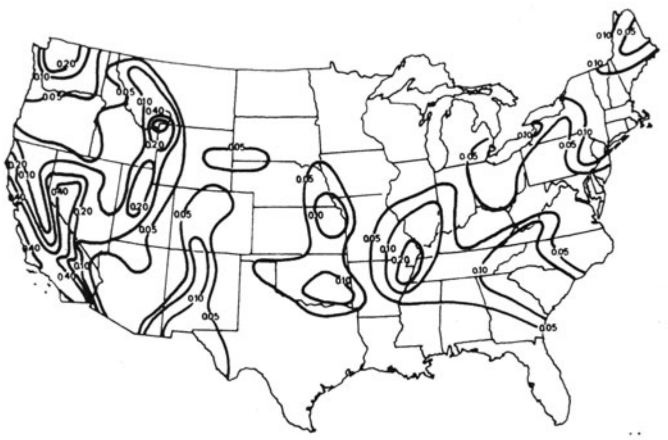


Fig. 8. Effective peak acceleration (EPA) for the lower contiguous United States. Contours represent EPA levels with a nonexceedance probability of between 80 and 95% during a 50-year period. Contours are for values of EPA in units of gravity. EPA is a modification of the peak-recorded instrumental acceleration designed to represent the energy content in the ground motion to which buildings respond. Although the estimate is somewhat more complex, it can be likened to the filtering out of high-frequency spikes of acceleration which structures do not react to because of their considerable inertia. (*America Society of Civil Engineers.*)

The basis of seismology is the observation and analysis of elastic energy as it propagates itself through the earth. When mechanical energy is released in a homogeneous earth, it is propagated outward in waves whose fronts are spherical and whose mechanism is alternating compression and rarefaction of the material through which they pass. These waves, called *P* waves, are physically analogous to the sound waves that spread outward from an explosion in air or water. The *P* wave travels at a rate of about 5.6 kilometers per second. It is the first wave to reach the surface. A longitudinal wave, the *P* wave tends to create a "push-pull" effect on rock particles as it passes.

The earth is not generally homogeneous, and though imperfectly elastic, it has rigidity or shear strength which is absent in air or water. This results in a second kind of wave action, in which the material through which the wave passes moves transversely to the direction of wave motion. These shear or *S* waves travel through solid material at a velocity a little more than half that of the *P* waves. The *S* wave causes the earth to move at a right angle to the direction of the wave.

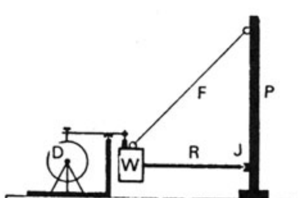


Fig. 9. Principle of a seismograph. *G*, ground; *P*, post set in the ground; *W*, weight; *R*, rigid support contacting the post with a free-moving sharp point at *J*; *F*, flexible wire; *D*, recording drum. Marker extends from *W* to *D*. An improved seismograph will employ an electrical pickup of pendulum movements. The instrument is placed on a concrete base which, in turn, is thoroughly anchored into bedrock so that when the instrument base moves it truly reflects earth movement and not extraneous vibrations.

In addition to the *P* and *S* types of waves (called body waves), there is also energy propagation along surfaces or interfaces. At the earth-air interface, waves similar to surface waves in water are propagated. These are called Rayleigh waves and cause, as in water, circular vertical motion of the material through which they pass, in the plane containing their direction of propagation. In solid material, surface waves like *S* waves also occur, the material moving horizontally in a plane transverse to the direction of motion of the wave. These waves are called Love waves (*L* waves). These waves usually are distinguishable only at great distances. *L* waves can cause the swaying of tall buildings as well as slight wave motions in bodies of water at great distances from the epicenter.

If the velocities of the different modes of wave propagation are known, deduction can be made, for example, of the distance between the earthquake and the observation station by measuring the time interval between the arrival of the faster and slower waves.

When earthquake waves move across an interface between earth materials with different elastic properties, their velocity, direction, and phase may be changed, and they will generally give rise to waves of other modes. For example, a *P* wave encountering an interface may give rise to a refracted and reflected *P* wave and a refracted and a reflected *S* wave. A wave may also be refracted even if it does not cross an interface, because of the change in pressure and density of the earth material as the wave penetrates more deeply into the earth.

Richter Scale. Earthquake size as determined by instruments is measured on a logarithmic scale called the Richter *magnitude* scale. In one variation of this scale, the very largest shocks have magnitudes slightly greater than 8.5. Energy in ergs is given empirically by $\log E = 11.4 + 1.5M$, where M is the magnitude. The measurements are based on records made on a standard type of seismograph a distance of 100 kilometers (62 miles) from the epicenter. Usually, seismograms from several different stations contribute to computing the magnitude of an earthquake. In most instances, of course, a station will not be the standard 100 kilometers from the epicenter. Thus, many records must be compared and complex conversion tables help to estimate the final figure. Fortunately, an experienced seismologist, within a few minutes, usually can estimate magnitude with reasonable accuracy from a record made at only one seismographic station. The logarithmic character of the Richter scale is often overlooked by lay people and news reporters. Obviously, an earthquake of 8.0 magnitude is not twice as powerful as one of 4.0 magnitude, but rather it is $10 \times 10 \times 10 \times 10$ times as powerful.

Modified Scales. In an attempt to provide a more meaningful and precise measure of earthquake energy, some seismologists are using a quantity called the *seismic moment*. This represents the product of the change in volumetric strain energy (the change in stress or stress drop multiplied by rigidity), the size of the fault rupture surface, and the actual displacement or throw of the fault. The seismic moment expressed in fundamental physical units becomes an incomprehensibly large number. For example, the seismic moment of the 1971 San Fernando earthquake ($M_L = 6.5$) was 1.4×10^{26} ergs. Research using a magnitude scale based upon seismic moment, developed by Kanamori (California Institute of Technology), has suggested revisions in the relative sizes of some historic earthquakes. The 1906 San Francisco earthquake would be reduced, while the 1960 Chilean earthquake, which may have been the largest event in this century, would be increased. Although seismic moment is more useful for the study of earthquakes, the complications brought about by the need to estimate three parame-

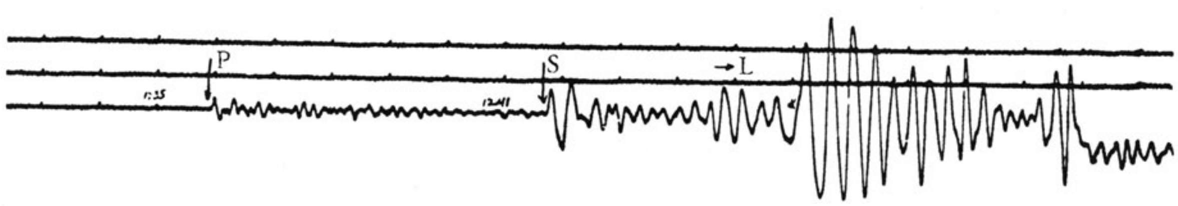


Fig. 10. Representative seismographic record (seismogram) of an earthquake. *P*, *S*, and *L* waves are described in text. (*University of California.*)

ters have delayed acceptance of this value in the development of standards for earthquake-resistant structures.

In comparing earthquakes, Kanamori first compared the estimated energy release with the magnitude of moderate-sized events. The extrapolation was then made to large events by taking the Gutenberg-Richter relationship between magnitude and energy, and using the energy release computed from the seismic moment. It should be noted that while the new scale provides a better measure of the relative size of events, the recorded magnitudes of events using the Richter scales remains unchanged.

Seismology in Petroleum and Minerals Exploration. Mineral explorers, notably in the oil and gas field, have found in recent years that the former, traditional two-dimensional seismic prospecting methods were inadequate for solving the three-dimensional problems encountered in petroleum exploration. In three-dimensional seismic methodology, the reflected wave field must be adequately sampled spatially to ensure proper imaging of the subsurface by means of a three-dimensional wave equation. The development of petroleum resources led to the use of seismic exploration techniques. The first field work in the United States began in the mid-1920s. After surface and less expensive means of exploration for gas or oil, seismographic methods may be used. Charges of explosives may be used, or large hydraulic vibrators, mounted on trucks, may be used to induce a shock wave series into the earth. Seismographs measure the reflected or refracted shock waves at carefully spaced locations. Interpretations of data from rock strata more than 6 km below the surface often provide accurate results. Offshore seismic searches are conducted by creating sharp sound wave pulses in the water which travel through the water and deeply into the rock formations below. Unfortunately, seismic information can only indicate subsurface conditions normally favorable to the generation and accumulation of gas or oil—not positively identify their presence. See Table 2.

TABLE 2. REPRESENTATIVE LONGITUDINAL SEISMIC WAVE VELOCITIES

Material	Velocity	
	Feet per Second	Meters per Second
Weathered surface material	500– 2,000	152– 610
Gravel, rubble, sand (dry)	1,500– 3,000	457– 914
Sand (wet)	2,000– 6,000	610–1829
Clay	3,000– 9,000	914–2743
Sandstone	6,000–13,000	1829–3962
Shale	7,000–14,000	2134–4267
Limestone	7,000–20,000	2134–6096
Granite	15,000–19,000	4572–5791
Metamorphic rocks	10,000–23,000	3048–7010
Glacial till (Saskatchewan)	5,000– 7,000	1524–2134
Ice	12,500	3810
Fresh water	4,700– 4,900	1433–1494
Seawater	4,800– 5,000	1463–1524

Data compiled from Jakosky and the Saskatchewan Research Council.

Earthquake Forecasting and Precursors

Particularly during the 1980s and early 1990s, the study of tectonic data for use in forecasting earthquake events has become more intensive. Some of the forecasting techniques, notably in the case of California, are discussed later in this article. In essence, historical information, combined with current seismic measurements, constitutes the basis for long- and short-range projections. These methodologies encompass the concepts of seismic gaps, dilatency, groupings of small earthquakes, and history of foreshocks. Paleoseismology has become an important part of investigative techniques. Less scientifically accredited are shorter-term precursors, such as the presence of radon gas in groundwater, presence of helium in fault outgas, the electrical conductivity/resistivity of rocks, meteorological data and soil tilt, gravitational pull of the sun and moon, and animal behavior just prior to tectonic events.

Paleoseismology. Related to finding seismic gaps, a number of specialists have gone back into earthquake history, not in terms of years or decades, but in terms of centuries to arrive at long-term patterns. Paleoseismology embraces the detailed study of landforms across fault zones, analysis of deformed layers of sediment in the walls of trenches excavated across active faults and the determination of the age of carbonaceous material found in the sediments, using radiometric techniques. K. E. Sieh (California Institute of Technology) and colleagues have been conducting surveys of this nature along the San Andreas fault, notably in the Pallet Creek area, since the mid-1970s. An initial analysis indicated a history of at least twelve large earthquakes in the last 1400 years, with an average recurrence interval between 140 and 150 years.

Tree rings also can be indicative of past tectonic action in a fault area because during such events some trees may be seriously affected. Roots growing across a fault may be severed, and vibration may shake off a tree's crown, actions that seriously affect tree growth. One group of scientists demonstrated by way of observing prior tree injuries, as obtained by analyzing tree rings, that a previously undetermined major earthquake occurred on the fault in 1812, thus shortening the seismic gap by about 44 years, instead of a previously assumed gap at that location of about 130 years. Such information can seriously alter the basis for models constructed to reflect assumed seismic gaps.

Historical Records—Seismic Gaps. Akin to Mendeleev's search for missing elements to fill out spaces in the Periodic Table of the Elements, seismologists today are making mid- and long-term predictions of earthquakes by identifying regions in which earthquakes have not occurred for a long period, but where theoretically they should have occurred. In other words, seismologists are selecting regions that are "overdue" for seismic action. The Tumaco, Colombia earthquake (December 12, 1979) of magnitude 8, the largest quake in northwestern South America since 1942, was predicted by Kelleher as early as 1972. This event filled a gap in the shallow seismic zone of northern Ecuador and southwestern Colombia, where no sizable earthquake had occurred since a quake of magnitude 8.7 happened in 1906. By carefully studying the region, Kelleher concluded that the direction of region during each prior significant earthquake in the general region was proceeding in a north to northeastern direction toward Tumaco.

The San Andreas fault in southern California presents a major seismic gap. The fault is not uniform. As reported by scientists of the U.S. Geological Survey, on one section of the fault in central California, fault creep (slow slippage) characterizes the displacement. In other places, the fault releases its stored elastic strain in characteristic steps no longer than a few centimeters. In other places, the characteristic slip increment is from 7 to 10 meters. Episodes of slippage of the latter type were responsible for the great earthquakes of 1857 and 1906.

It has been suggested that, if averaged over a sufficiently long period of time, the sum of the various slippages or displacements along the fault should equal the displacement between the two underlying plates. Thus, the possibility for long-term predictions based upon what might be called "slip budget." By applying this method in conjunction with Reid's elastic-rebound theory, scientists have formulated long-range predictions of earthquake potential, namely, that large earthquakes will occur somewhere along the fault at intervals of from 50 to 200 years.

In their excellent report of the Tumaco earthquake, Herd et al. (1981) observe that history suggests that another series of large, shallow-focus earthquakes along the Ecuador-Colombia coastline may begin in this century near Esmeraldas. The 1942 Esmeraldas earthquake (magnitude, 7.9), the first of the latest series of northward-progressing earthquakes in the Ecuador-Colombia seismic gap, followed the 1906 earthquake by only 36 years. In the years since 1942, there have been no large-magnitude earthquakes near Esmeraldas. Another shallow-focus earthquake could recur in this new seismic gap at Esmeraldas before the balance of the 1906 seismic gap is filled near Buenaventura.

Another example of the "seismic gap" forecasting technique is the Alaskan earthquake of February 1979, which occurred in a sparsely populated part of the Alaskan coast, some 400 kilometers (248 miles)

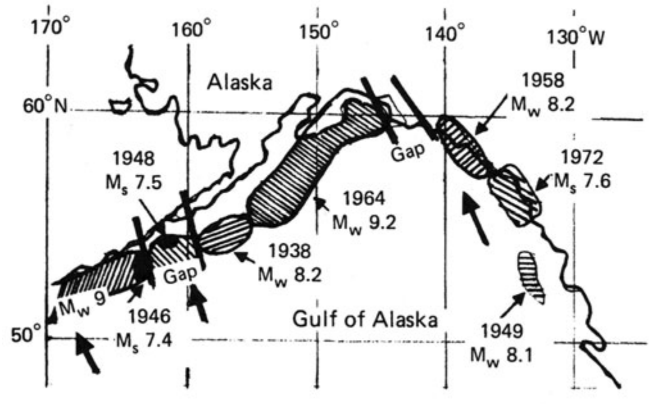


Fig. 11. Two gaps in the Alaska-Aleutian seismic zone. Top gap was noted in 1971. The 1979 earthquake ruptured only the eastern half of the gap. (*After McCann, Lamont-Doherty Geological Observatory.*)

east of Anchorage. As shown by Fig. 11, the earthquake filled one of at least two seismic gaps, as forecasted as early as 1971. Another large quake in the remainder of the gap, which extends almost to Valdez, is expected within the next decade or so.

The earthquake forecast diagram shown in Fig. 12, based largely on the seismic gap theory, illustrates how well the USGS forecasted the 1989 Loma Prieta earthquake.

Groupings of Smaller Earthquakes—Foreshocks. As of the early 1990s, the search for correlations between the groupings of smaller earthquakes and foreshocks and larger subsequent earthquakes is receiving much attention. The theory is that nondestructive earthquakes and foreshocks indicate a build-up of stress that must eventually be released in the form of a main shock of considerable magnitude. However, in some earthquake-prone regions, such as California, scientists have been frustrated for the last few years by the absence of small quakes from which to obtain data. This area of interest has currently displaced the attention which scientists gave a few years ago to dilatancy-related effects.

Although both events exhibited foreshocks, neither the Izu-Oshima-kinkai earthquake in Japan of 1978 nor the Adak event in the Aleutians could be interpreted with sufficient reliability to warrant a short-term prediction. In retrospect, however, pre-main event data could be used to forecast similar events should precipitating steps be similar, a broad assumption.

Not all earthquakes have foreshocks of at least moderate magnitude. Of 160 earthquakes of magnitude 7 or greater included in a survey by the California Institute of Technology, only half exhibited foreshocks of magnitude 4 or greater. Thus, it is obvious that lesser earthquakes must be included in possible foreshock groupings in an effort to expand the data base and yet remain above the random earth noise level. Clusters of events of magnitude 3 and 4 are now included by some investigators. Also, foreshocks useful to forecasting purposes may be separated by much greater time spans than previously considered. For example, it appears that there was a subtle pattern of seismicity over a ten-year period in the case of the San Fernando, California earthquake of 1971.

Investigators at several institutes have proposed the concept of a "slow earthquake," for which the crustal movement is about one hundred times slower than that of a normal quake. Such a slow quake would not produce higher frequency seismic signals recorded by standard seismographs and thus would go undetected. Nevertheless, such an earthquake could transmit stress from one place to another. In the opinion of one investigator, slow earthquakes may be responsible for a sudden increase of stress to critical levels within only a few days or weeks prior to a final failure. If indeed slow earthquakes do occur, this would greatly complicate the problem of making long-term predictions.

Seismic amplitude measurements made by the U.S. Geological Survey in 1978 suggest that foreshocks sometimes may have different focal mechanisms than aftershocks. The ratio of the amplitudes of *P* and *S* waves from the foreshocks and aftershocks of three California earthquakes (Oroville, Galway Lake, and Briones Hills) during the 1975–1977 period showed a characteristic change at the time of the main events. As this ratio is extremely sensitive to small changes in the orientation of the fault plane, a small systematic change in stress or fault configuration in the source region may be inferred. The success of the Chinese in predicting the Haicheng earthquake (February 1975) was based upon the use of long-, intermediate- and short-term precursors. As pointed out by Lindh et al. (1978), the imminent prediction was

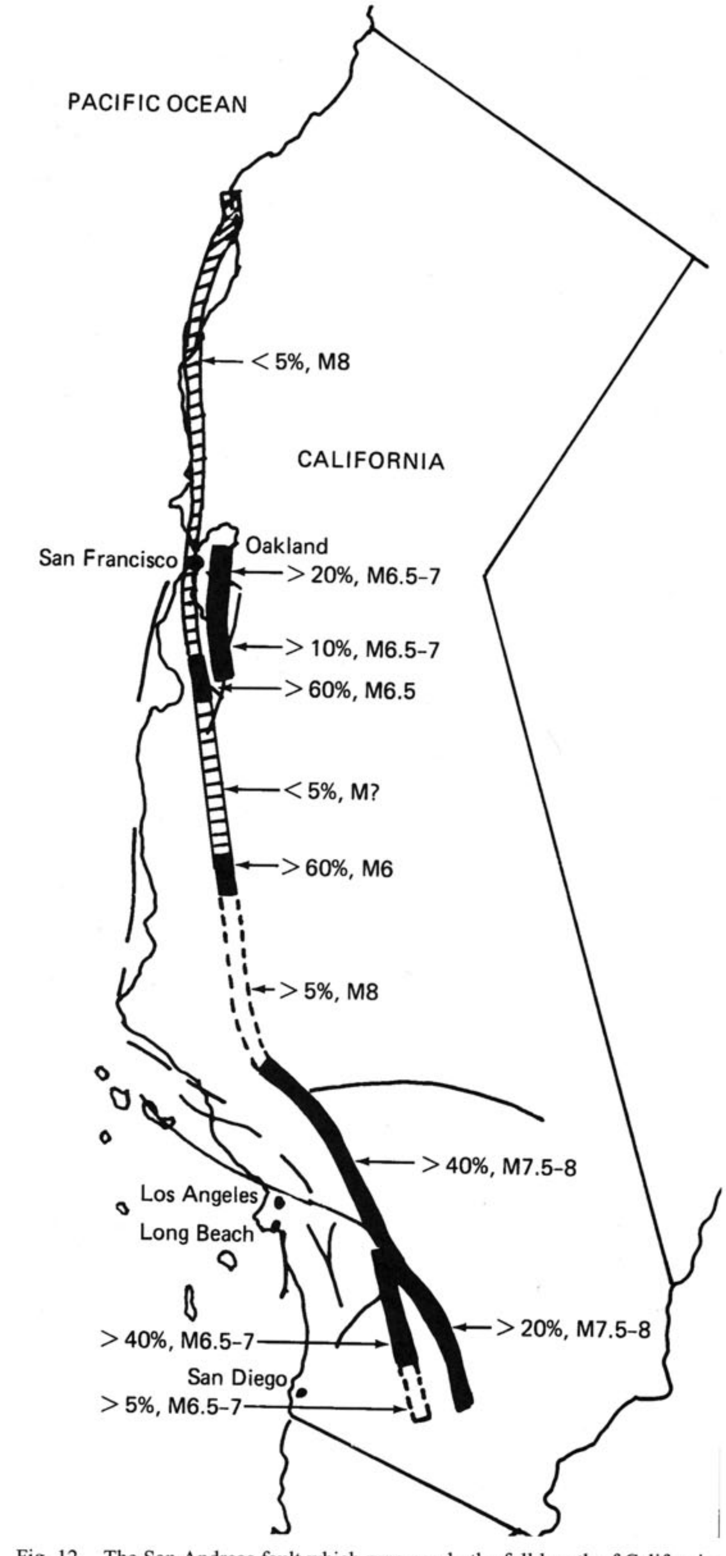


Fig. 12. The San Andreas fault which runs nearly the full length of California. Probability of a great earthquake occurring between the present time and the year 2015 is expressed as a greater than (>), or less than (<) percentage. If earthquake occurs along designated portions of the fault, the expected magnitude (*M*) is given after the percentage probability figure. Other single black lines show locations of additional principal faults in the state. (*Source: U.S. Geological Survey.*)

based partly on the recognition of a swarm of small-to-moderate earthquakes near Haicheng as a potential foreshock sequence. However, foreshocks do not always occur; estimates of their frequency vary widely and they are usually recognizable as foreshocks only in hindsight. To use foreshocks in a predictive mode, one must distinguish them from background seismicity and earthquake swarms.

Dilatancy Theory. H. R. Reid (Johns Hopkins University) a number of years ago proposed that earthquakes are generated by the sudden slip and elastic rebound of crustal blocks bordering a fault. The crust could bend under strain, then snap and straighten out, thus releasing the energy of the strain. A model of this theory involves many microscopic events preceding an earthquake and many observations over a considerable time span are needed to fill out the dilatancy model. The application of the theory has now been combined with the concept of plate tectonics.

A precursor that is directly implied by dilatancy is *crustal uplift*. Such uplift may be detected by geodetic or tilt measurements. Where a dilatant zone is sufficiently large as well as shallow, dilatancy could produce a crustal uplift of several centimeters. In the case of an earthquake of magnitude 6 (Odaigahara, Japan, 1960), five tiltmeters at different localities up to 100 kilometers from the epicenter detected precursory movements. Six months prior to the earthquake, rapid tilting in the direction away from the epicenter began to be recorded at two stations, 40 and 100 kilometers from the epicenter. Simultaneously a strainmeter at one of the sites began to show anomalously high rates of extension. These activities continued for 3 months, after which they stopped. However, a station farther to the south began tilting. One month prior to the earthquake, all three stations began tilting in the opposite direction and two additional stations began to tilt rapidly. The data indicate that different stages of dilatancy occurred at different times at different localities. Both the magnitude and the duration of the tilts are explicable on the basis of the dilatancy model. The reverse of tilt just prior to the earthquake indicates some type of short-term precursor, perhaps due to a slight compaction of the dilatant volume.

Presence of Radon Gas. As early as 1966, the year of the Tashkent earthquake, analysis of waters from a deep well in the hypocentral region indicated a variation in the radon content prior to the earthquake. Radon has a half-life of only 3.8 days, and its lifetime diffusion distance is only several centimeters. Thus, its increased concentration in the Tashkent well prior to the earthquake and its larger aftershocks could have been caused by two means: (1) an increase in the surface area of rock in the epicentral region due to cracking, or (2) by an increase in the flow rate of pore water. Both of these effects are predicted by the dilatancy model. In the case of the Izu-Oshima-kinkai, Japan earthquake of 1978, precursory changes in the radon concentration of groundwater were observed. The distance from the epicenter to a continuous radonmonitoring station at Nakaizu was about 25 kilometers (15.5 miles). A sudden drop and a subsequent increase in the radon concentration recorded 5 days prior to the earthquake were significant. The size of the spikelike change was about 15%. After the earthquake, a remarkable increase in the radon concentration occurred.

In a 1980 paper, five Japanese scientists made the following observations: "Although our understanding of the whole mechanism of radon emission remains unclear, the observed precursory changes must somehow reflect the deformation or damage, or both, to the artesian layer caused by the action of stress release or stress accumulation. Even though we did not predict the occurrence of the Izu-Oshima-kinkai earthquake, our efforts were directed toward obtaining data on reliable premonitory changes through observation, a first step toward earthquake prediction. We conclude that the measurement of the radon concentration in groundwater at carefully chosen wells should offer definitive information on the likelihood of an earthquake."

In California, in 1979, radon gas concentration was measured in two wells along the San Jacinto fault, on the northern margin of the Los Angeles basin. A marked increase in concentration was noted during the period of minor seismic activity. More data are needed to make a judgment, according to some west coast geophysicists.

Helium "Spots." Japanese scientists have found that measurements of variations in the degassing rate from a fault can be a good way of inferring the most recent time of displacement along a fault, including displacements, such as creep and stick slippage. Helium is believed to be one of the most ideal elements to measure for this purpose because of its mobility, chemical inertness, and low abundance in the atmosphere. Using helium as a geochemical tracer, the investigators aimed at a qualitative assessment of an active fault related to the occurrence of an earthquake. It should be pointed out that helium surveys already have been attempted for prospecting for natural gases, ores, and active geothermal areas. Application of a similar technique to active fault zones may provide useful information for earthquake prediction.

Electrical Conductivity. A measurement associated with the concept of dilatancy is a significant reduction of the electrical resistivity of rock and soil prior to an earthquake as a result of the influx of water into the dilatant region. In the Garm region of Russia, strong correlations were shown some years ago between minima in electrical resistivity and the occurrence of earthquakes. These findings were supported by studies of the electrical resistivity of rocks in the laboratory.

Meteorological Data and Soil Tilt. A popular instrumentation strategy for earthquake prediction is to deploy large numbers of relatively inexpensive tiltmeters, strainmeters, creepmeters, and other geophysical instruments. Most of these instruments are implanted at shallow depths in soil regimes near active faults. In some cases, the sites are within the fault zone. Scientists at the U.S. Geological Survey (Menlo Park, California), over a two-year period, studied the numerical relation between earthquake activity, temperature, and rainfall, as well as tiltmeter response. Their conclusions were that, if the data contain premonitory earthquake signals, they are buried in local meteorological noise. Separating an earthquake anomaly from the response to surface phenomena becomes more difficult as the earthquake anomaly lead time approaches the rise time of the soil to weather and seasonal variations.

Animal Behavior. The reports of animal behavior as a precursor of an imminent earthquake are taken seriously by some researchers and not really ruled out by any investigators. The following excerpt from the Haicheng Earthquake Study Delegation (China, 1979) is of interest: "Some instances noted at this time (before the earthquake) were of snakes being found frozen on the road...geese flying, chickens refusing to enter their coops, pigs rooting at their fence, cows breaking their halters and escaping, and goats as well as cows being unusually restless. Rats appeared to behave as though drunk. Three well-trained police dogs howled, refused to obey commands, and kept their noses close to the ground as though sniffing." In China, numerous non-scientifically trained persons serve as observers along with scientific coordinators. The content of the foregoing excerpt may be better understood after perusing the Shapley (1976) reference listed.

Most investigations have shown that peculiar animal behavior occurs during or after, rather than before an earthquake. If animal precursors are to be considered seriously, there must be some scientific explanation of the sensing mechanisms possessed by birds, farm animals, pets, etc., that make them more perceptive than humans, or even of scientific instruments. One possible explanation is the superior sensitivity of some animals to perceive very low, negligible (from a human standpoint) vibrations of earthquakes of magnitude 2 and under. These may arrive as very gentle foreshocks of a large quake. Only a very nearby seismometer would pick up such vibrations. If felt by only one instrument in California's dense network of seismometers, for example, the data would not be considered valid. There is also the possibility that animals will detect very low, rumbling booms with a frequency of 50–70 Hz, similar to faint thunder. Humans can detect such noises, but some species, such as pigeons, are far more sensitive to frequencies below 50 Hz. Again, such mild vibrations could arrive ahead of a large shock.

Seismologists have noticed the barking of dogs during the gentle foreshocks of weak quakes as well as the mild aftershocks. Some of these effects were observed during the Willits, California earthquake of 1977. However, an extensive monitoring activity, from which several reports were received, failed to provide convincing documentation of peculiar animal behavior prior to the Coyote Lake, California quake of November 1979. At that time, there were over 1200 volunteer observers in the region.

To date, no details have been worked out (except possibly in China) to include animals in a detection network for providing earthquake warnings sufficiently in advance to be of practical value.

Peculiar Earthquake Phenomena. Rumbling noises are definitely associated with some earthquakes and are not a figment of the imagination. At least since the time of Benjamin Franklin, there have been reports of a peculiar light "which illuminated the night sky" during an earthquake. Somewhat parallel to the animal behavior reports, the matter of an earthquake light persists. Thompson (1978) reports that mysterious phenomena, including earthquake light, have been chronicled since ancient times. Geologists noted a brightening of the sky the night before a major earthquake in Turkey (1975). In Japan, sightings and photographs have documented bright night skies before an earthquake. In one modern theory, it is proposed that hydrocarbons convert to methane, which is capable of surviving the high pressures and temperatures deep within the earth. It is further proposed that methane escaping through faults in the land surface catches up small grains of dust or dirt that collide, creating sparks that could ignite the methane and cause the flames and explosions sometimes reported prior to and during earthquakes. It has also been suggested that methane, combining with sulfur to form hydrogen sulfide may also be responsible for the observed uneasiness of animals.

A troubling aspect of the 1989 Loma Prieta earthquake was the absence of premonitory tremors. However, a serendipitous finding of radio signals discovered by Stanford University electrical engineers may prove important. Their monitoring of signals was not whatsoever related to an interest in earthquakes. For a number of years the group has been interested in very low frequency (VLF) radio waves in connection with the development of a communications medium for use with submarines. Thus, during the time frame of the Loma Prieta event, the scientists had been monitoring VLF waves. A few years prior, the Stanford engineers had moved their antenna from the Stanford campus to a location that would not be affected by background radiation. This location happened to be Corralitos, located only $6\frac{1}{2}$ km from the epicenter of the Loma Prieta quake. After cleaning up debris from the quakes, the investigators found that VLF waves had increased dramatically about 3 hours prior to the quake. They noted that the effect, involving ultralow (ULF) waves, definitely was associated with the Loma Prieta event. The researchers offer the hypothesis that: increased stress on rocks along the fault line generated a surge of piezoelectricity, or pressure-induced current, that in turn produced the radio waves. The fact that only the lowest-frequency (.01 Hz) waves were detected may imply that the prequake activity occurred well beneath the earth's surface. However, not all geophysicists agree with this concept.

Hydrothermal Precursors. Some researchers at The Carnegie Institution, Department of Terrestrial Magnetism, Washington, D.C., (see Silver and Valette-Silver reference) summarized the finding of a recent study. "During the period 1973 to 1991, the interval between eruptions from a periodic geyser in northern California exhibited precursory variations 1 to 3 days before the largest earthquakes within a 250-km radius of the geyser. These include the magnitude 7.1 Loma Prieta earthquake for which a similar preseismic signal was recorded by a strainmeter located halfway between the geyser and the earthquake. These data show that at least some earthquakes possess the observable precursors, one of the prerequisites for successful earthquake prediction. All three earthquakes were further than 130 km from the geyser, suggesting that precursors might be more easily found around rather than within the ultimate rupture zone of large California earthquakes." Hydrothermal data currently are used as part of earthquake forecasting in the United States, China, and Russia. In addition to Loma Prieta, the other two earthquakes mentioned were the Oroville earthquake of 1975 and the Morgan Hill event of 1984, both located in California.

Regional Tectonics and Paleotectonics

Because the interface between the earth's crust and the mantle is not consistently uniform under the continents and oceans, surface and near-surface phenomena (mainly earthquakes and volcanism) exhibit different characteristics from one geographic location to the next. There are parts of the earth's surface that are relatively quiet—that is, where major earthquakes have not occurred during the recent past and where volcanoes are not present. One must appreciate, of course, that the geologic time scale is measured in terms of millions of years. By contrast, there are very active, sometimes called hot areas where the present and recent past tectonic activity has been great. In fact, a large portion of the earth's tectonic activity presently is occurring in a region sometimes called the "Ring of Fire," which geographically includes the continental borders on either side of the Pacific Ocean. See also **Volcano** in this encyclopedia.

South and Central Western United States. The earth's crust is actively deforming in several regions. Because the San Francisco earthquake of 1906 occurred along the San Andreas fault, it has been the focus of attention by geophysicists for several decades.

San Andreas Fault. This fault, which lies along the border defined by the Pacific plate and the North American plate, is located along the California coastline, commencing under the Pacific Ocean some distance north and west of San Francisco, with landfall several miles north of San Francisco. Because horizontal slip of this fault under the San Francisco Bay area was determined as the cause of the San Francisco earthquake of 1906, much geophysical research has been conducted pertaining to the San Andreas fault, which is comprised of numerous segments that occur over a distance of some 600 miles (960 km), thus affecting southern California as well as further north.

Coachella Valley Segment of the San Andreas Fault. As recently as June 1992, the Landers earthquake (M = 7.4) occurred in California's Mojave Desert, 240 km east of Los Angeles. This further alerted scientists to the probability of a potential large-magnitude earthquake in the Los Angeles area. The Landers event emphasized the clustering of earthquakes very near to the Coachella Valley segment of the San Andreas fault. These included prior events at Big Bear, Palm Springs, and Joshua Tree. Geophysicists are becoming increasingly concerned with so many quakes occurring so close together on the northern end of the segment. One investigator has observed that an M = 7.5 earthquake on the Coachella segment alone would not cause the "Big One," but if the rupture broke into the next segment north and west, it could trigger an earthquake of M = 7.8 just at the edge of heavily populated San Bernardino. Rupture of this segment in line would continue past Los Angeles proper and could produce a quake of M = 8.

This view conflicts with the observations of the next paragraph and demonstrates the complexity and difficulty of predicting earthquakes based mainly on statistical evidence.

Recent studies have shown that the San Andreas fault is not necessarily suspect in all major earthquakes in California that may occur in the future. Some scientists, for example, have observed that the next large quake (the so-called "Big One") may not be attributed to the San Andreas fault, particularly if it occurs in southern California.

A model constructed in the 1960s indicated that the Pacific and North American plates have been moving relative to each other at a rate of 60 mm per year over the past 2 to 3 million years. During the 1970s, researchers, assisted by much improved instrumentation, indicated that movement along the San Andreas fault did not account in full for the relative movements of the two plates. Three causes were suggested to explain the discrepancy:

1. The early plate-movement rate was too high;
2. The missing motion may have included spreading of the Great Basin[3] and the Range Province of Utah and Nevada; or
3. Deformation in coastal California.

Scientists essentially concur that there is a combination of the three aforementioned factors. This, in essence, removes some of the onus from the San Andreas fault and some of the ancillary faults located near it.

These findings have been confirmed through the study of horst and graben structures and scarps, which are common to areas that have been compressed or stretched as the result of distant plate activity. In other words, plate interactions are not confined just to plate boundary interactions, but may spread for long distances along the breadth and length of a plate. Thus, in addition to faults, large areas of crust can be deformed and contribute to the total absorption of energy from underlying plate action. These energy-absorbing areas, so to speak, can reduce the total energy borne by surface fractures and thus contribute to area stability. On the other hand, these absorbing areas can become saturated and, like fractures, can become focal points for earthquakes. Such

[3]A region of the western United States located between the Sierra Nevada and the Wasatch mountains, including parts of Nevada, California, Idaho, Utah, Wyoming, and Oregon.

events may be variously referred to as fold, deep, or hidden earthquakes.

By way of perspective, one must appreciate that the theory of plate tectonics was not generally accepted by most geophysicists until the 1960s. Consequently, new hypotheses pertaining to details frequently are proposed.

As observed by Segall and Lisowski (see reference), "the horizontal displacements accompanying the 1906 San Francisco earthquake and the 1989 Loma Prieta earthquake are computed from geodetic survey measurements."[4] The 1906 earthquake displacement field is entirely consistent with the right-lateral strike slip on the San Andreas fault. In contrast, the 1989 Loma Prieta earthquake exhibited sub-equal components of strike slip and reverse faulting. These results, together with other seismic and geological data, may indicate that the two earthquakes occurred on two different fault planes.

Loma Prieta Earthquake, 1989. Much of the information that follows is a condensation of a report by the U.S. Geological Survey staff, released in January 1990.

The first major earthquake on the San Andreas fault since 1906 fulfilled a long-term forecast for its rupture in the southern Santa Cruz Mountains. Severe damage occurred at distances of up to 100 km from the epicenter in areas over ground known to be hazardous in strong earthquakes. Stronger earthquakes will someday strike closer to the urban centers in the United States, most of which also contain hazardous ground. The Loma Prieta earthquake demonstrated that meaningful predictions can be made of potential damage patterns and that, at least in well-studied areas, long-term forecasts can be made of future earthquake locations and magnitudes. Such forecasts can serve as a basis for action to reduce the threat that major earthquakes pose to the United States.

The surface wave magnitude, M, of the earthquake was 7.1. The quake was felt as far away as Los Angeles to the south and Reno, Nevada, to the east. Confirmed fatalities numbered 62 persons, injuries were 3757, with 12,000 persons left homeless. Property losses were estimated at $6 billion. Even though a few other major earthquakes have occurred in the United States since the Great San Francisco Earthquake in 1906, Loma Prieta produced dramatic nationwide reactions and concern, these resulting from excellent and extensive television coverage. Prior major U.S. earthquakes since 1906 included the Kern County, California, earthquake (July 1952, $M = 7.5$) and the Great Alaskan Earthquake (March 1964, $M = 9.2$).

The Loma Prieta earthquake ruptured a segment of the San Andreas fault in the Santa Cruz Mountains that had been recognized as early as 1983 as having a high probability of rupture within the following few decades. In a USGS study of 1988, this segment was assigned the highest probability for producing an M 6.5 to 7 earthquake of any California fault segment north of the Los Angeles metropolitan area.

Some Unfortunate Parallels with 1985 Mexico City Earthquake. Just as much damage and as many casualties were caused by the Michoacan earthquake (official name) in Mexico City, some 350 km distant, as by the Loma Prieta quake, which reached parts of San Francisco and Oakland, much of its damage caused by effects on ground fills and the phenomenon of *liquefaction*. Inadequate freeway design was a cause of severe damage. Damage assessment of Loma Prieta revealed unreinforced brick masonry and structures having soft, open-ground floors that have inadequate resistance to shear deformation induced by strong earthquake shaking. It is interesting to note that modern buildings in nearby San Jose generally held up well during the earthquake. In the central San Francisco Bay Area, structures known to be of earthquake vulnerability were damaged. San Francisco's marina district was built over landfill and hence suffered from liquefaction; gas lines and water lines were broken, thus furnishing the fuel for a disastrous fire in that district and delaying ample water supplies to put the fires out. The Loma Prieta earthquake reinforced the observations of experts to the effect, "The amount of damage produced by an earthquake depends chiefly on the geological character of the ground. Where the surface is of solid rock, the shock usually produces little damage, whereas in structures on landfills, great violence is manifested."

The Loma Prieta earthquake generated landslides throughout a region of about 14,000 square kilometers. Away from the zone of surface fractures, landslides were most numerous in the Santa Cruz mountains and consisted of rock falls, rock slides, and debris slides.

Fortunate Short, Sharp Shock. The strong shaking associated with the Loma Prieta earthquake persisted only for some 6 to 10 seconds, as contrasted with the Armenia earthquake, 1988, when shaking went on for a period of 30 seconds. At an Earthquake Engineering Research Institute (Univ. of California, Berkeley) meeting held a few months after the Loma Prieta earthquake, one scientist observed that the 40–50-km rupture of the fault segment that caused the earthquake began at a central point, traveling outward in two directions at once. Had the fault ruptured unidirectionally, strong shaking could have persisted for 20 to 30 seconds. Experts observe that a number of borderline structures that survived had been stressed nearly to their limits and, had the shaking continued for just a little longer, many large buildings would have collapsed. In several instances, large diagnonal cracks developed in some of the taller buildings, suggesting that just a little more pounding would have brought them down. Some structural engineers, in analyzing the damage rendered to the Oakland Bay Bridge, also express some good fortune in the manner in which both the earthquake forces and the bridge performed. Two short stretches of roadway in the eastern third of the bridge collapsed, with the loss of only one life. The roadway segments that fell were the upper and lower roadways at pier E9 of the bridge. The fallen bridge segments were the upper and lower roadways at pier E9, midway across the eastern stretch of the bridge. Subsequent damage assessment of the bridge indicated that inertial forces within the bridge pulled the trusses and roadway east of pier E9 toward Oakland by about 7 inches (18 cm) relative to the main portions of the bridge. Each of the numerous piers supporting the bridge use forty 1-inch diameter bolts to secure the roadways with each pier. Fortunately, in the opinion of a structural engineer, if the initial shocks of the earthquake had not ruptured these particular bolts, the forces would have been transferred to pier E9 per se, with much greater damage being done to the bridge. In retrospect, it appears that the failed bolts on one pier served as a protective energy-relief device, much as a safety valve on a boiler. Several hundred more cars and their passengers could have been thrown into the bay had the bolts not sheared. A map and more detail can be found in the Barinaga reference listed.

Liquefaction. This phenomenon occurred in the 1987 Superstition Hills, California earthquake and in the Loma Prieta earthquake, notably in the San Francisco marina area built upon landfill. As defined by Holzer, Youd, and Hanks, "Seismically induced liquefaction involves the loss of static shearing resistance of saturated, relatively loose, sandy deposits due to a tendency to closer packing of the constituent grains dynamically driven by seismic shear waves. If pore fluid in the liquefying layer cannot escape, this reduction in pore volume causes porewater pressure to increase. Liquefaction is generally thought to occur when pore pressure approaches lithostatic. Common surface manifestations of liquefaction include fountains of water laden with sediment and ground failure.

Fold Earthquakes in California. Some California earthquakes identified as *fold earthquakes* include:

1. Coalinga earthquake, 1983, $M = 6.3$, causing no deaths, but demolishing 75% of unreinforced structures.
2. Kettleman Hills earthquake, 1985, $M = 6.1$, located in a remote area, but felt over a wide area.
3. Whittier Narrows earthquake, 1987, $M = 6.0$. This event struck within the populous Los Angeles basin and, as pointed out by Stein and Yeats, "The Whittier Narrows quake was only one-tenth the size of the Coalinga event, but it caused 10 times the damage, that is, $150 million and taking 8 lives."

[4]Horizontal displacements during the 1989 Loma Prieta earthquake have been measured with a variety of techniques. Most information has come from laser electronic distance measurements (EDM). Changes in distance constrain horizontal displacements up to rigid body translations and rotations of the network. Global positioning system (GPS) measurements between Loma Prieta and several other locations, including Eagle, Mount Hamilton, and Fort Ord, constrain the relative displacement vectors between these sites and thus the rotational component of the displacement field. Displacement of the Fort Ord site relative to stations remote from the epicentral region have been determined by very long baseline interferometry (VLBI), constraining the transitional components of the displacement field.

In none of these earthquakes did underlying faults slip-cut the surface of the earth. The earthquakes occurred on young anticlines less than several million years old. In each case, the fold heightened measurably during the earthquake. From this pattern of performance, it can be reasoned that not only can young anticlines be sites of earthquakes, but also that the folds actually are created by tectonic activity—that is, by a series of earthquakes over time, but with little if any evidence at the surface. It was determined that the Coalinga earthquake differed from surface-fault quakes in its pattern of aftershocks. Unlike surface-fault earthquakes, in which the aftershocks tend to be aligned along the plane of the fault, in the Coalinga event the aftershocks were distributed more diffusely, above and below the fault plane. This could suggest that the Coalinga anticline involves a number of faults. See also encyclopedia entry on **Anticline**.

The Whittier Narrows earthquake lifted its associated anticline, the Santa Monica mountains fold. Some geologists suggest that the Whittier Narrows event has been but one of many similar events that have taken place along a blind fault that runs for 150 km underneath the California coastline. Also, it is suggested that the Point Magu earthquake, 1973, M = 5.6, also may have been a fold earthquake. Another anticline, known as the Ventura Avenue anticline, is located on the southern California coast. Although no major earthquake has been registered along this anticline, paleotectonics records indicate that it may be related to other California anticlines previously described.

Geophysicists have noted that anticline fold earthquakes rupture comparatively slowly. Rupture time for the Kettleman Hills earthquake was estimated at 16 seconds—that is, about four times longer than a typical surface-faulting earthquake. The rupture time at Coalinga compared with most surface quakes, but it was observed that the seismic activity took much longer to die out. Some investigators note that these characteristics may be related to the slower release of pressure from fluids contained in the rocks. These observations may reflect a scenario of slow, more prolonged earthquakes. Questions still unanswered: Do anticline faults undergo steady and continuous slip, or do they grow intermittently through earthquakes equal to or larger than the Whittier Narrows event? Stein and Yeats have observed, "It is incumbent on seismologists to distinguish between the two competing explanations; their consequences are dramatically different. If earthquakes larger than the 1987 event are possible beneath the Santa Monica mountains fold, then Los Angeles' greatest earthquake threat may not come from a future M8 shock on the San Andreas fault, 50 km to the north, but from a smaller earthquake originating under downtown Los Angeles." Investigators are probing this possibility.

Some clues may be derived from further studies of fold, deep, and hidden earthquakes that have occurred in other parts of the world. Earthquakes between M7 and M7.8 have occurred on blind faults in Argentina, Canada, northern India, Japan, and New Zealand, with a high potential for such earthquakes in the Balkans, Chile, Iran, Italy, Pakistan, and Taiwan.

Brief mention should be made of the El Asnam, Algeria 1980 M7.3 shock, which killed 3,500 people in three North African cities, and the Armenian 1988 (M = 6.4) earthquake, which claimed a minimum of 25,000 lives. The aftershock pattern of the latter earthquake was diffuse. The area of the Armenian event had been mapped previously by the former USSR Academy of Science. After the event, observations were made by a *Landsat* satellite.

The Armenian earthquake has been described as a "worst case scenario." The city of Spitak is less than 5 km from the fault. Just 4 minutes after the mainshock of M = 6.8, a major aftershock of M = 5.8 collapsed most of the buildings that had been weakened by the mainshock. The majority of people were killed by the collapse of old, weakened buildings that incorporated no shock-resistant structural design. Even some of the more recent high-rise buildings were severely damaged.

Some so-called very deep focus earthquakes (focal depths of 50–1200 km) occur from causes not previously described and poorly understood. Subduction zones appear to be the setting for such events. The most destructive earthquake of this type to occur within recent times occurred in Bucharest, Rumania, 1977 (M = 7.2). The focal depth was estimated at 150 km. Another deep-focus earthquake occurred 650 km under Colombia, M = 7.6. A number of investigators, notably Wadati and Benioff, have develoed hypotheses as to the active physics and mechanism of such quakes. These concepts are summarized by Frohlich (see reference).

Very Short Term Alerting Systems in California. Most earthquake prediction systems, as previously described, are geared to the long term—that is, expectations at best are expressed in terms of years. Such predictions are useful toward increasing the public's awareness of how necessary it is to support preparative measures, such as reinforcing structures, setting up agencies for coordinating earthquake emergencies, and providing practical safety precautions to individuals, much as is done in connection with the threats of tornadoes and hurricanes. Conventional earthquake prediction procedures obviously do not serve those immediate actions that can be taken if warning can be given in just a few minutes or even seconds prior to a shock.

During the past decade, considerable research has been directed toward *very short term* warnings, even though initially this may seem scientifically implausible. However, by way of sophisticated and strategically placed earthquake instrumentation systems, warnings essentially in terms of "real time" may be possible. Applications, however, appear to be applicable only to tectonically limited areas. Nevertheless, a combination of such measurement systems could serve wider warning areas. In 1993, considerable research along these lines is under way.

Seismic Versus Radio Waves. Radio waves travel considerably faster than seismic waves; thus, over a distance, site instrumentation data can telegraph ahead seismic information from a given site. For example, most damaging seismic waves propagate out from a fault rupture at a rate of about 3 km per second, obviously slow in comparison with the radio transmission of instrumentation data measured at a given site. Within a distance of some 32 km (20 mi), there is a 10-second time difference of arrival time. Over a distance of 320 km (300 mi), there would be a time differential of about 200 seconds (3 minutes). By integrating such segments of an active fault, however, the warning or alerting time span could be increased.

The Parkfield Experiment. Locations along the San Andreas fault and the Parkfield segment in particular have produced reasonably uniform seismic patterns over a period of years. But, of course, there always have been fairly wide margins of error in terms of predicting exact times of occurrence. At Parkfield, past earthquakes have displayed very similar characteristics (seismic patterns). These are typical of a so-called *periodic* quake, based upon the seismic gap theory. Thus, geophysicists regard the Parkfield fault as a "model."

Since the late 1880s, magnitude 6 quakes have occurred at Parkfield on an average of one every 22 years. Quakes occurred in 1881, 1901, 1922, 1934, and 1966. These data have been confirmed even from far-distant seismic centers, such as one in the Netherlands, in addition to centers in the United States. Parkfield, a sparsely populated community (less than 50 people), was not selected because of possible local damage or loss of life, but because of its past reliable quake record. A quick calculation shows that another quake should have occurred in 1988 and thus, as of early 1992, is overdue, but so was the one in 1966. To the geophysicist, similarity of seismic data is far more important than exact timing.

The Parkfield experiment was established in 1985 and since then many millions of dollars have been invested by the USGS, the state of California, and universities worldwide. The overall aim is to develop a warning system in terms of weeks, days, hours, or even seconds of a magnitude 6 earthquake.

A "real-time" warning system has been established that provides four levels of alert:

Level D. This is the lowest level and indicates that some *unusual* seismic or slip activity has occurred on the fault segment, but that there is less than 1% chance of a magnitude 6 earthquake within 72 hours. Approximately 50 Level D alerts had been issued as of early 1990.

Level C. This occurs when (a) numerous arrays of instruments located along the fault segment record low levels of activity, or (b) one array of instruments records a high level of activity. This indicates that there is less than a 5% change of a magnitude 6 earthquake. About 20 Level C alerts were issued between 1987 and early 1990.

Level B. This represents an 11–37% probability of a magnitude 6 quake occurring within the next 72 hours.

Level A. This represents that the probability of an earthquake occurring within a very short time span is 1 in 2.

No Level B or A alerts have been issued since the system was established.

Arguments that favor a real-time warning system, even if the warning period is very short, include turning off gas supplies and thus preventing quake-started fires, putting emergency agencies on highest alert, warning hospitals, alerting vehicular traffic, securing radioative materials, and, depending upon the time interval available, interrupt public radio and television programs to warn the public. As one scientist stated, a person can crawl under a desk in just a few seconds.

A similar system in Japan alerts high-speed rail trains to come to a stop as quickly as possible without causing injury to passengers.

Parkfield, unfortunately, was too far from the October 1989 Loma Prieta earthquake to participate, but if a short warning could have been issued from another station along the San Andreas fault, perhaps gas could have been turned off and thus not available to fuel the fire that occurred in the marina area of San Francisco.

Seismic Instrumentation at Parkfield. Some sets of state-of-the-art and more sophisticated instrumentation installed along the Parkfield segment include a strong-motion array, a distributed strong-motion array, a differential strong-motion array, a surface geology effects array, a dense strong-motion array, a liquefaction array, and a pipeline experiment, among other systems.

As reported by Allen Lindh (USGS), strain measurements make up the backbone of monitoring instrumentation. These sensors determine the deformation of rock at a single point. This requires a precision of about 1 part per billion. This precision is impossible to achieve at the surface, and thus the strain gages are installed about 300 meters (1000 feet) down into the earth. A device referred to as a "two-color laser geodimeter" is used to detect any distance changes between hilltops up to 10 km (6+ mi) apart with a precision of 1 mm. Such measurements indicate any movement of crustal blocks on the earth's surface. Any change of "slip" pattern is indicative of a forthcoming quake. Different arrays of seismometers are used to measure almost infinitesimal motions of the earth's surface and record acoustics and other kinds of energy radiated by an earthquake. Such data are computer analyzed to estimate locations and magnitudes of small quakes within 3 to 5 minutes. These data are helpful for determining possible foreshocks, which could signal an imminent large quake.

Additional instrumentation at Parkfield includes magnetometers, which detect changes in the earth's magnetic field, and creep meters, which measure surface slip on the fault. Deep-water wells also are monitored to determine possible rock deformation. Under test is a telluric current monitoring array, which senses changes in resistivity to electrical currents in the earth.

Communication and Computer System. Data from the aforementioned instruments are transmitted by microwave or satellite to a center at Menlo Park, California. For this experimental system, USGS scientists are availabe around the clock to analyze data whenever the computer finds a situation corresponding with the previously mentioned levels of alert.

Generally based upon the assumption that the Parkfield experiment will perform well, planning for a much more ambitious and sophisticated warning system involving all important segments of the San Andreas fault is well underway. Already, of course, there are scores of fault measuring systems installed along the fault. These are to be upgraded to include the most sophisticated instrumentation. The basis of an integrated system, of course, would be extensive analysis of past performance of many of the segments of the San Andreas fault. Added to the integrated system at the start would be data gathering stations along an essentially connected fault running from north of San Francisco to considerably southeastward of Los Angeles. Somewhat further into the future, other major faults, such as Hayward to mention only one, could be added to the system.

The overall technology, under development at Massachusetts Institute of Technology, essentially would be based on an artificial intelligence (AI) program in a computer linked to accelerometers and other instrumentation along the fault line. In an "if this, then" fashion, the computer, operating at a very high data processing rate, would determine the magnitude of a quake and the areas most likely to sustain damage. Warnings then would be sent automatically by radio or telephone directly to regions at risk, at which time warning sirens and other emergency operations would commence. The program is being written in LISP artificial intelligence language and will run on a standard 386 chip-based personal computer.

Instrumentation that will feed data to the master center will be located approximately 5.6 km (0.6 mi) apart. The accelerometers to be used will be about .03 cubic meters (1 cubic foot) in size and mounted on concrete pads, ideally set in bedrock to provide the most precise information.

Northwestern United States. Exercises in paleoseismology have revealed that a large earthquake occurred during the eighth or ninth century A.D. almost directly under present-day Seattle, Washington. Investigators describe the Juan de Fuca plate as being subducted beneath the North American plate, moving in an east-northeast direction. Geodetic data concerning deformations in northeastern Washington reveal that strain is accmulating. The average rate of strain accumulation in the Olympic network has been estimated through the use of triangulation surveys.

Moderate earthquakes occurred under Puget Sound in 1949 and 1965, the latter causing an estimated $12 million in damage. Since that quake, a tight code on building construction has been implemented. Radioactive carbon dating and tree ring counting (dendrochronolog) currently are underway, out of which models of past tectonic activity in the area will be constructed. This research is underway by the U.S. Geological Survey staff and the University of Washington. There has been insufficient time, as of 1993, to assess the hazard posed by the Seattle fault. Officials also are concerned that an earthquake in the area could create tsunamis in Puget Sound.

Central and Eastern North America. By normal tectonic standards, those regions of America east of the Rocky Mountains are considered "quiet." However, a few of the greatest earthquakes in terms of magnitude throughout the world have occurred in continental crust—not near the boundary of two plates, but at distances well within the length and breadth of a given plate. Such events sometimes are referred to as "interplate" earthquakes. Their focal points generally are deep; their frequency of occurrence usually is in terms of scores of years or centuries, in contrast with the comparatively short terms of repetition as experienced in plate-boundary regions.[5]

Of the 15 recorded intraplate earthquakes occurring on stable crust that has been stretched and thinned over the past 250 million years, 6 occured on part of the North American plate (eastern North America): New Madrid Seismic Zone (3 times); Charleston, South Carolina; Grand Banks, Newfoundland, 1929 (M = 7.4); and Baffin Bay, West Greenland, 1933 (M = 7.7).

All of these earthquakes occurred well within the boundaries of the North American plate, recalling that the eastern boundary of this plate extends well eastward, to the Mid-Atlantic ridge.

The events occurred where crust has been stretched and thinned over the past 250 million years.[6]

New Madrid Seismic Zone. The New Madrid Fault is centrally located in this zone. This fault lies in a slightly north easterly direction, extending from Memphis, Tennessee, on the south and nearly reaching Paducah, Kentucky, at its northern terminus. Historically affected seis-

[5]An exception to this generalization was the occurrence of three events in the New Madrid seismic zone. One event occurred in 1811 (M = 8.2), and two events occurred in 1812 (M = 8.1 and M = 8.2).

[6]The remaining (worldwide) recorded intraplate events on stable crust (See Johnson/Kasten reference) were:

EURASIAN PLATE:
Basel, Switzerland, 1356 (M = 7.4); Taiwan Straits, 1604 (M = 7.7); Haiman Island, 1605 (M = 7.7); Portugal, 1858 (M = 7.1); Nanai, China, 1918 (M = 7.4).

INDIAN-AUSTRALIAN PLATE:
Kutch, India, 1819 (M = 7.8); Exmouth Plateau, Australia, 1906 (M = 7.6); South Tasman Rise, 1951 (M = 7.0).

AFRICAN PLATE:
Libya, 1936 (M = 7.1).

mic areas have included western Tennessee and Kentucky, eastern Missouri and Arkansas, southern Indiana and Illinois, and northern Mississippi.

Newspapers of the day described the 1812 event in impressive, but unconfirmed terms: "It toppled chimneys in distant Cincinnati, Ohio, rang church bells in far-off Boston, Massachusetts, awakened President James Madison at the White House, and Thomas Jefferson at Monticello in Virginia."

As reported by Liu, Zeback, and Segall (see reference), first-and second-order triangulation networks were established in the area as early as 1929. In the 1950s, a much wider area was surveyed with the use of second-order triangulation. Although crustal deformation can be measured with repeated angle measurements from triangulation data alone, the aforementioned investigators, associated with Stanford University and the USGS concluded that there were insufficient repeated angles in the region to compute strain. Consequently, a new survey of many of the triangulation stations with the global positioning system (GPS) was implemented in 1991. "This made it possible to determine whether detectable crustal strain had accumulated during the past 35 to 40 years." Rapid crustal strain accumulation since the 1950s was detected. A tentative conclusion was drawn to the effect that the observed strain rates were due to post-seismic, lower-crustal flow in response to the 1811–1812 events, rather than strain accumulation associated with impending earthquakes.

Mistaken Warning of a December 1990 Event. In the fall of 1990, Dr. Iben Browning, described as a self-taught climatologist and who previously had made outstanding forecasts of a few prior seismic events and thus had established a form of track record,[7] but one that ranged from skepticism and luck to some degree of recognition among a few professionals and notably in the nonscientific community, declared that a large earthquake would occur along the New Madrid fault on December 3, 1990. The event *did not* occur and, as of early 1993, no such event has occurred. However, at least one school district in the area closed schools on December 3 and 4. Missouri and Arkansas National Guards also were put on alert. The entire affair became a major media event. Although very interesting, more detail cannot be given here. The reader is referred to the Kerr, August 9, 1991, reference listed.

Charleston, South Carolina Earthquake. The largest earthquake recorded in the eastern coastal areas of the United States occurred near Charleston, South Carolina, in August 1886. Because this event (M = 7.0+) affected a populous area, it is generally considered to have been the most destructive of earthquakes occurring in the United States during the nineteenth century. It has been estimated that a similar event in the same location today would result in several thousand fatalities and property damage of $400 million or more.

Amick and Gelinas (Ebasco Services Incorporated, West Meadows, North Carolina) have conducted research to assess the probability of future earthquakes in the region. Because Charleston is quite distant from an active plate boundary and there is an absence of any convincing paleoseismic evidence of faulting or other characteristic physical phenomena normally associated with such tectonic events, such an assessment indeed is difficult and carries an aura of mystery to be solved.

In their excellent report (see reference), Amick and Gelinas summarize their observations to date, "The special distribution of seismically induced liquefaction features discovered along the Atlantic seaboard suggests that during the last 2000 to 5000 years, large earthquakes (body magnitude, $m_b \geq 5.8 + 0.4$) in this region may have been restricted exclusively to South Carolina. Paleoliquefaction evidence for six large prehistoric earthquakes was discovered there. At least five of these past events originated in the Charleston, South Carolina, area, the locale of a magnitude 7+, including the 1886 event. During the past two millennia, large events have occurred about every 500 to 600 years."

[7]Dr. Browning, who apparently considered the gravitational pull of the sun and moon for his forecasts, previously had close-coupled forecasts in connection with the eruption of Mount St. Helens in 1980, the Mexico City quake event of 1985, and the Novado del Ruiz eruption in Colombia in 1985.

Additional Reading

Adams, J.: "Paleoseismology: A Search for Ancient Earthquakes in Puget Sound," *Science*, 1592 (December 4, 1992).

Amick, D., and R. Gelinas: "The Search for Evidence of Large Prehistoric Earthquakes Along the Atlantic Seaboard," *Science*, 655 (February 8, 1991).

Amos, J. L., and D. Jeffery: "Fossils: Annals of Life Written in Rock," *Natl. Geographic*, 182 (August 1985).

Anderson, D. L.: "Composition of the Earth," *Science*, 243 (January 20, 1989).

Anderson, D. L., Tanimoti, T., and Yu-shen Zhang: "Plate Tectonics and Hotspots: The Third Dimension," *Science*, 1645 (June 19, 1992).

Appenzeller, T.: "Deeply Moved (Earth's Core)," *Sci. Amer.*, 24 (April 1988).

Appenzeller, T.: "Coming Down in Sheets (A Model Mantle Recreates Motifs of Plate Tectonics," *Sci. Amer.*, 17 (August 1989).

Appenzeller, T.: "Just a Veneer (Upper Continental Crust)," *Sci. Amer.*, 26 (November 1989).

Appenzeller, T.: "Roving Stones," *Sci. Amer.*, 19 (February 1990).

Apperson, K. D.: "Stress Fields of the Overriding Plate at Convergent Margins and Beneath Active Volcanic Arcs," *Science*, 670 (November 1, 1991).

Audley-Charles, M. G., and A. Hallman, Editors: "Gondwana and Tethys," Oxford Univ. Press, New York, 1988.

Badash, L.: "The Age-of-the-Earth Debate," *Sci. Amer.*, 90 (August 1989).

Bak, P., and K. Chen: "Self-Organized Criticality," *Sci. Amer.*, 46 (January 1991).

Baker, E. T.: "Megaplumes," *Oceanus*, 84 (Winter 1991/1992).

Bally, A. W., and A. R. Palmers, Editors: "The Geology of North America: An Overview," Geological Society of America, Boulder, Colorado, 1989.

Barinaga, M.: "Loma Prieta: Saved by a Short, Sharp Shock," *Science*, 1390 (December 15, 1989).

Baringa, M.: "Quick Fix for Freeways," *Science*, 31 (January 5, 1990).

Beardsley, T.: "Aftershocks," *Sci. Amer.*, 14 (January 1990).

Bell, D. R., and G. R. Rossman: "Water in Earth's Mantle: The Role of Nominally Anhydrous Minerals," *Science*, 1391 (March 13, 1992).

Ben-Zion, Y., and P. Malin: "San Andreas Fault Zone Head Waves Near Parkfield, California," *Science*, 1592 (March 29, 1991).

Bercovici, D., Schubert, G., and G. A. Glatzmaier: "Three-Dimensional Spherical Models of Convection in the Earth's Mantle," *Science*, 960 (May 26, 1989).

Bergman, E. A.: "Mid-Ocean Rise Seismicity," *Oceanus*, 60 (Winter 1991/1992).

Bethke, C. M., et al.: "Supercomputer Analysis of Sedimentary Basins," *Science*, 239 (January 15, 1988).

Bilham, R., and P. Bodin: "Fault Zone Connectivity: Slip Rates on Faults in the San Francisco Bay Area, California," *Science*, 281 (October 9, 1992).

Bird, P.: "Formation of the Rocky Mountains, Western United States: A Continuum Computer Model," *Science*, 1501 (May 25, 1988).

Blackman, D., and T. Stroh: "Ridge: Cooperative Studies of Mid-Ocean Ridges," *Oceanus*, 21 (Winter 1991/1992).

Bloxham, J., and D. Gubbins: "The Evolution of the Earth's Magnetic Field," *Sci. Amer.*, 68 (December 1989).

Bohlen, S. R., and K. Mezger: "Origin of Granulite Terranes and the Formation of the Lowermost Continental Crust," *Science*, 326 (April 21, 1989).

Bolt, B. A.: "Inside the Earth: Evidence from Earthquakes," W. H. Freeman, Salt Lake City, Utah, 1982.

Bolt, B. A.: "Balance of Risks and Benefits in Preparation for Earthquakes," *Science*, 169 (January 11, 1991).

Boraiko, A.: "Earthquake in New Mexico," *Nat'l. Geographic*, 654 (May 1986).

Boraiko, A.: "The Rise and Fall of Buildings—A Primer for Survival in Quake City," *Nat'l. Geographic*, 664 (May 1986).

Bowen, D. Q.: "The Last 130,000 Years," *The University of Wales Review*, 39 (Spring 1989).

Bryan, W. B.: "From Pillow Lava to Sheet Flow Evolution of Deep-Sea Volcanology," *Oceanus*, 42 (Winter 1991/1992).

Bucknam, R. C., Hemphill-Haley, E., and E. B. Leopold: "Abrupt Uplift Within the Past 1700 Years at Southern Puget Sound, Washington," *Science*, 1611 (December 4, 1992).

Cann, J.: "Onions and Leaks: Magma of Mid-Ocean Ridges," *Oceanus*, 36 (Winter 1991/1992).

Carroll, J.: "Tremors of Fear (Browning and the New Madrid Fault)," *San Francisco Chronicle*, 1: Also reproduced in *The Courier Jrnl.* (Louisville, Kentucky), 1 (July 31, 1990).

Cathles, L. M., III: "Scales and Effects of Fluid Flow in the Upper Crust," *Science*, 323 (April 20, 1990).

Chen, I.: "Pre-Quake Quirks: Searching for Predictors," *Sci. News.*, 231 (October 13, 1990).

Cloud, P.: "Oasis in Space. Earth History from the Beginning," Norton, New York, 1988.

Coppersmith, K. J., et al.: "Methods for Assessing Maximum Earthquakes in the Eastern United States," *EPRI Report EPRI rp-2556-12,* Electric Power Research Institute, Palo Alto, California, 1990.

Cullen, V.: "On Mid-Ocean Ridges," *Oceanus* (Winter 1991/1992).

Davies, G. F.: "Plates and Plumes: Dynamos of the Earth's Mantle," *Science*, 493 (July 24, 1992).

Dillon, W. P., et al.: "Geology of the Caribbean," *Oceanus*, 42 (Winter 1987/88).
Dziewonski, A. M., and J. H. Woodhouse: "Global Images of the Earth's Interior," *Science*, 37 (April 3, 1987).
Edmond, J. M.: "Himalayan Tectonics, Weathering Processes, and the Strontium Isotope Record in Marine Limestones," *Science*, 1594 (December 4, 1992).
Epstein, A. W.: "Woody Witnesses (Tree Rings and Earthquakes)," *Sci. Amer*., 24 (September 1988).
Ford, J. P., et al.: "Faults in the Mojave Desert, California, as Revealed on Enhanced *Landsat* Images," *Science*, 1000 (May 25, 1990).
Frohlich, C.: "Kiyoo Wadati and Early Research on Deep Focus Earthquakes: Introduction to Special Section on Deep and Intermediate Focus Earthquakes," *J. Geophysical Research*, Vol. 92, No. B13, 13777–13788 (December 10, 1987).
Frolich, C.: "Deep Earthquakes," *Sci. Amer*., 48 (January 1989).
Fukao, Y,: "Seismic Tomogram of the Earth's Mantle: Geodynamic Implications," *Science*, 625 (October 23, 1992).
Fumal, T. E., et al.: "A 100-Year Average Recurrence Interval for the San Andreas Fault at Wrightwood, California," *Science*, 199 (January 8, 1993).
Germanovich, L. N., and R. P. Lowell: "Percolation Theory, Thermoelasticity, and Discrete Hydrothermal Venting in the Earth's Crust" *Science*, 1564 (March 20, 1992).
Gerster, G., and G. Kieffer: "Africa Gives Slow Birth to a New Ocean," *Telegraph Sunday Magazine (London)*, Issue No. 127 (1978).
Gordon, R. G., and S. Stein: "Global Tectonics and Space Geodesy," *Science*, 333 (April 17, 1992).
Gore, R., and J. A. Sugar: "Our Restless Planet Earth," *Natl. Geographic*, 142 (August 1985).
Graf, W. L., Editor: "Geomorphic Systems of North America," Geological Society of America, Boulder, Colorado, 1987.
Gurnis, M.: "Ridge Spreading, Subduction, and Sea Level Fluctuations," *Science*, 970 (November 16, 1990).
Gurnis, M.: "Rapid Continental Subsidence Following the Initiation and Evolution of Subduction," *Science*, 1556 (March 20, 1992).
Hallam, A.: "Alfred Wegener and the Hypothesis of Continental Drift," in *Scientific Genius and Creativity*, W. H. Freeman, New York, 1987.
Hanks, T. C.: "Small Earthquakes: Tectonic Forces," *Science*, 1430 (June 5, 1992).
Hanson, R. B.: "Planetary Fluids," *Science*, 281 (April 20, 1990).
Harrison, T. M., et al.: "Raising Tibet," *Science*, 1663 (March 27, 1992).
Hart, S. H., et al.: "Mantle Plumes and Entrainment: Isotopic Evidence," *Science*, 517 (April 24, 1992).
Haukasson, E., et al.: "The 1987 Whittier Narrows Earthquake in Los Angeles Metropolitan Area, California," *Science*, 1409 (March 18, 1988).
Herd, D. G., et al.: "The Great Tumaco, Colombia Earthquake of 12 December 1979," *Science*, **211**, **441** (1981).
Hide, R., and J. O. Dickey: "Earth's Variable Rotation," *Science*, 629 (August 9, 1991).
Hill, R. I., et al.: "Mantle Plumes and Continental Tectonics," *Science*, 186 (April 10, 1992).
Hoffman, K. A.: "Ancient Magnetic Reversal: Clues to the Geodynamo," *Sci. Amer.*, 76 (May 1988).
Hoffman, P. F.: "Did the Breakout of Laurentia Turn Gondwanaland Inside-Out?" *Science*, 1409 (June 7, 1991).
Holden, C.: "U.S. Drilling Takes New Tack," *Science*, 158 (January 11, 1991).
Holzer, T. L., Youd, T. L, and T. C. Hanks: "Dynamics of Liquefaction During the 1987 Superstition Hills, California Earthquake," *Science*, 56 (April 7, 1989).
Horgan, J.: "Broadcast Warning," *Sci. Amer*., 26 (March 1990).
Jacoby, G. C., Jr., Sheppard, P. R., and K. E. Sieh: "Irregular Recurrence of Large Earthquakes Along the San Andreas Fault: Evidence from Trees," *Science*, 196 (July 8, 1988).
Jacoby, G. C., Williams, P. L., and B. M. Buckley: "Tree Ring Correlation between Prehistoric Landslides and Abrupt Tectonic Events in Seattle, Washington," *Science*, 1621 (December 4, 1992).
Jaumé, S. C., and L. R. Sykes: "Changes in State of Stress on the San Andreas Fault Resulting from the California Earthquake Sequence of April to June 1992," *Science*, 1325 (November 29, 1992).
Johnston, A. C., and L. R. Kanter: "Earthquakes in Stable Continental Crust," *Sci. Amer*., 68 (March 1990).
Jordan, T. H., and J. B. Minster: "Measuring Crustal Deformation in the American West," *Sci. Amer*., 48 (August 1988).
Karson, J. A.: "Tectonics of Slow-Spreading Ridges," *Oceanus*, 51 (Winter 1991/1992).
Kerr, R. A.: "Making Deep Earthquakes in the Laboratory," *Science*, 30 (January 5, 1990).
Kerr, R. A.: "Puzzling Out the Tectonic Plates," *Science*, 808 (February 16, 1990).
Kerr, R. A.: "The Lessons of Dr. Browning," *Science*, 622 (August 9, 1991).
Kerr, R. A.: "Deep Rocks Stir the Mantle Pot," *Science*, 783 (May 10, 1991).
Kirby, S. H., Durham, W. B., and L. A. Stern: "Mantle Phase Changes and Deep-Earthquake Faulting in Subducting Lithosphere," *Science*, 216 (April 12, 1991).
King, G. C. P., and C. Vita-Finzi: "Active Folding in the Algerian Earthquake of 10 October 1980," *Nature*, 22 (July 2, 1981).
Knittle, E., and R. Jeanloz: "Earth's Core-Mantle Boundary: Results of Experiments at High Pressures and Temperatures," *Science*, 1438 (March 22, 1991).
Krogstad, E. J. et al.: "Plate Tectonics 2.5 Billion Years Ago: Evidence at Kolar, South India," *Science*, 1337 (March 10, 1989).
Kroner, A., and P. W. Layer: "Crust Formation and Plate Motion in the Early Archean," *Science*, 1405 (June 5, 1992).
Lambo, L., Zoback, M. D., and Paul Segall: "Rapid Interplate Strain Accumulation in the New Madrid Seismic Zone," *Science*, 1666 (September 18, 1992).
Lay, T., et al.: "Studies of the Earth's Deep Interior: Goals and Trends," *Physics Today*, 44 (October 1990).
Libbrecht, K. G., and M. F. Woodard: "Advances in Helioseismology," *Science*, 152 (July 12, 1991).
Lin, J.: "The Segmented Mid-Atlantic Ridge," *Oceanus*, 11 (Winter 1991/1992).
Lindh, A., Fuis, G., and C. Mantis: "Seismic Amplitude Measurements Suggest Foreshocks Have Different Focal Mechanisms than Aftershocks," *Science*, **201,** 56 (1978).
Lindh, A. G.: "Earthquake Prediction Comes of Age," *Technology Review (MIT)*, 42 (February/March 1990).
Liu, L., Zoback, M. D., and P. Segall: "Rapid Intraplate Strain Accumulation in the New Madrid Seismic Zone," *Science*, 1686 (Settember 15, 1992).
Lonsdale, P., and C. Small: "Ridges and Rises: A Global View," *Oceanus*, 26 (Winter 1991/1992).
Lowenstein, T. K., Spencer, R. J., and Z. Pengxi: "Origin of Ancient Potash Evaporites: Clues from the Modern Nonmarine Qaidam Basin of Western China," *Science*, 1090 (September 8, 1989).
Lutz, R. A.: "The Biology of Deep-Sea Vents and Seeps," *Oceanus*, 75 (Winter 1991/1992).
Macdonald, K. C.: "Mid-Ocean Ridges: The Quest for Order," *Oceanus*, 9 (Winter 1991/1992).
Malin, P. E., et al.: "Microearthquake Imaging of the Parkfield Asperity," *Science*, 557 (May 1, 1989).
Malin, P. E., and M. G. Alvarez: "Stress Diffusion Along the San Andreas Fault at Parkfield, California," *Science*, 1005 (May 15, 1992).
McCammon, C.: "Effect of Pressure on the Composition of the Lower Mantle End Member Fe_xO," *Science*, 66 (January 1, 1993).
McCann, D. M., and P.D. Jackson: "Seismic Imaging of the Rock Mass," *Review (Univ. of Wales)*, 21 (Autumn 1988).
McNutt, M. K., and A. V. Judge: "The Superswell and Mantle Dynamics Beneath the South Pacific," *Science*, 969 (May 25, 1990).
Meade, C., and R. Jeanloz: "Deep-Focus Earthquakes and Recycling of Water into the Earth's Mantle," *Science*, 68 (April 5, 1991).
Menard, H. W.: "The Ocean of Truth: A Personal History of Global Tectonics," Princeton Univ. Press, Lawrenceville, New Jersey, 1986.
Merrill, R. T., and P. L. McFadden: "Paleomagnetism and The Nature of the Geodynamo," *Science*, 145 (April 20, 1990).
Michael, A. J., and D. Eberhart-Phillips: "Relations Among Fault Behavior, Subsurface Geology, and Three-Dimensional Velocity Models," *Science*, 651 (August 9, 1991).
Minster, J. B., and T. H. Jordan: "Vector Constraints on Western U. S. Deformation from Space Geodesy, Neotectonics, and Plate Motions," *J. Geophysical Research*, Vol. 92, No. B6, 4798–4808 (May 10, 1987).
Monastersky, R.: "Birth of a Subduction Zone," *Science News*, 396 (December 16, 1989).
Monastersky, R.: "Rattling the Northwest," *Sci. News*, 104 (February 17, 1990).
Monastersky, R.: "Loma Prieta's Long-Distance Punch," *Science News*, 251 (April 21, 1990).
Monastersky, R.: "Tibet's Tectonic Escape Act," *Sci. News*, 2 (July 14, 1990).
Murphy, J. B., and R. D. Nance: "Supercontinent Model for the Contrasting Character of Late Proterozoic Orogenic Belts," *Geology*, Vol. 19, No. 5, 469–472 (May 1991).
Murphy, J. B., and R. D. Nance: "Mountain Belts and the Supercontinent Cycle," *Sci, Amer*., 84 (April 1992).
Mutter, J. C.: "Seismic Images of Plate Boundaries," *Sci. Amer*., 66 (February 1986).
Mutter, J. C.: "Seismic Imaging of Sea-Floor Spreading," *Science*, 1442 (November 27, 1992).
Nance, R. D., Worsley, T. R., and J. B. Moody: "The Supercontinent Cycle," *Sci. Amer*., 72 (July 1988).
NCAR: "Network Newsletter," published monthly by the Environmental and Societal Impacts Group, National Center for Atmospheric Research, Boulder, Colorado (January 1992).
Newsom, H. E., et al.: "Origins of the Earth," Oxford Univ. Press, New York, 1990.
Newsom, H. E., and K. W. W. Sims: "Core Formation During Early Accretion of the Earth," *Science*, 926 (May 17, 1991).

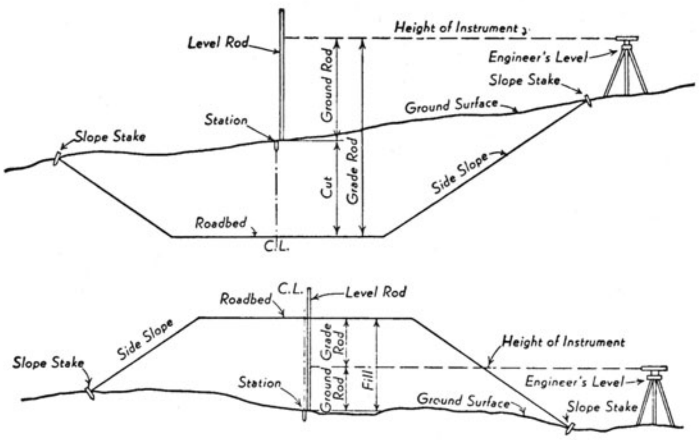
Ng, Max K.-F., Leblond, P. H., and T. S. Murty: "Simulation of Tsunamis from Great Earthquakes on the Cascadia Subduction Zone," *Science*, 1248 (November 30, 1990).
Nishenko, S. P., and G. A. Bollinger: "Forecasting Damaging Earthquakes in the Central and Eastern United States," *Science*, 1412 (September 21, 1990).
NRC: "Real-Time Earthquake Monitoring," National Research Council, Washington, D.C., 1991.
Parsons, T., and G. A. Thompson: "The Role of Magma Overpressure in Suppressing Earthquakes and Topography: Worldwide Examples," *Science*, 1139 (September 20, 1991).
Peacock, S. M.: "Fluid Processes in Subduction Zones," *Science*, 329 (April 20, 1990).
Powell, C. S.: "Peering Inward," *Sci. Amer.*, 100 (June 1991).
Reasenberg, P. A., and L. M. Jones: "Earthquake Hazard After a Mainshock in California," *Science*, 1173 (March 3, 1989).
Reasenberg, P. A., and R. W. Simpson: "Response of Regional Seismicity to the Static Stress Change Produced by the Loma Prieta Earthquake," *Science*, 1687 (March 27, 1992).
Reitter, L.: "Earthquake Hazard Analysis," Columbia Univ. Press, New York, 1991.
Richards, M. A., et al.: "Flood Basalts and Hot-Spot Tracks: Plume Heads and Tails," *Science*, 103 (October 6, 1989).
Ruddiman, W. F., and J. E. Kutzbach: "Forcing of Late Cenozoic Northern Hemisphere Climate by Plateau Uplift in Southern Asia and the American West," *J. of Geophysical Research*, Vol. 94, No. D15, 18409–18427 (December 1989).
Ruddiman, W. F., and J. E. Kutzbach: "Plateau Uplift and Climatic Change," *Sci. Amer.*, 66 (March 1991).
Sacks, P. E., and D. T. Secor, Jr.: "Kinematics of Late Paleozoic Continental Collision Between Laurentia and Gondwana," *Science*, 1702 (December 21, 1990).
Sauatter, V., Haggerty, S. E., and S. Field: "Ultradeep Ultramafic Xenoliths: Petrological Evidence from the Transition Zone," *Science*, 827 (May 10, 1991).
Savage, J. C., and M. Lisowski: "Strain Measurements and the Potential for a Great Subduction Earthquake Off the Coast of Washington," *Science*, 101 (April 5, 1991).
Schlesinger, W. H., et al.: "Biological Feedbacks in Global Desertification," *Science*, 1043 (March 2, 1990).
Schlesinger, W. H.: "Biogeochemistry," Academic Press, San Diego, California, 1991.
Scholz, C. H.: "The Mechanics of Earthquakes and Faulting," Cambridge Univ. Press, New York, 1990.
Schouten, H., and J. Whitehead: "Ridge Segmentation: A Possible Mechanism," *Oceanus*, 19 (Winter 1991/1992).
Segall, P., and M. Lisowski: "Surface Displacements in the 1906 San Francisco and 1989 Loma Prieta Earthquakes," *Science*, 1241 (November 30, 1990).
Seidler, N. M., and L. Kohl: "Origin of Earth and Life," *Natl. Geographic*, 152 (August 1985).
Shapley, D.: "The Chinese Earthquake," *Science*, **193**, 656 (1976).
Silver, P. G., and N. J. Valette-Silver: "Detection of Hydrothermal Precursors to Large Northern California Earthquakes," *Science*, 1363 (September 4, 1992).
Siuru, B.: "Mission to Planet Earth," *Sensors*, 48 (April 1992).
Smith, R. J.: "U.S., Soviets Share Seismic Posts," *Science*, 807 (August 25, 1989).
Smylie, D. W.: "The Inner Core Translational Triplet and the Density Near Earth's Center," *Science*, 1678 (March 27, 1992).
Soren, D.: "The Day the World Ended at Kourion—Reconstructing an Ancient Earthquake," *Nat'l. Geographic*, 30 (July 1988).
Sornette, A., and D. Sornette: "Self-Organized Criticality and Earthquakes," *Europhysics Letter*, Vol. 9, No. 3, 197–202 (1989).
Staff: "Off California, Even the Islands are Moving," *Nat'l. Geographic*, News Section (October 1992).
Stager, C., and C. Johns: "Africa's Great Rift," *Nat'l. Geographic*, 2 (May 1990).
Stauffer, N. W.: "Rooftop Laboratory," *Technology Review (MIT)*, 10 (April 1988).
Stebbins, J. F., and I. Farnan: "Nuclear Magnetic Resonance Spectroscopy in the Earth Sciences: Structure and Dynamics," *Science, 257* (July 21, 1989).
Stein, R. S., and R. S. Yeats: "Hidden Earthquakes," *Sci. Amer.*, 48–57 (June 1989).
Stein, R. S., King, G. C. P., and J. Lin: "Changes in Failure Stress on the Southern San Andreas Fault System Caused by the 1992 Magnitude = 7.4 Landers Earthquake," *Science*, 1328 (November 20, 1992).
Stix, G.: "Future Shock," *Sci. Amer.*, 111 (January 1991).
Stix, G.: "Finding Fault: Can Seismologists Predict Earthquakes?" *Sci. Amer.*, 48 (December 1992).
Tishchenko, A. A.: "TSVET-1 Colorimeter," *Aviatsiya i kosmonavitka*, 44, Issue No. 1 (1984).
Tishchenko, A. A.: "Space Coloristics," *Human Engineering*, 3, Issue No. 5 (1987).
Tivey, M. K.: "Hydrothermal Vent Systems," *Oceanus*, 68 (Winter 1991/1992).
Toomey, D. R.: "Tomographic Imaging of Spreading Centers," *Oceanus*, 92 (Winter 1991/1992).
U.S. Geological Survey Staff: "The Loma Prieta, California, Earthquake: An Anticipated Event," *Science*, 286 (January 19, 1990).
Van Dam, L.: "Reducing Disasters During Earthquakes," *Technology Review (MIT)*, 12 (February/March 1990).
Van der Voo, R.: "Paleozoic Paleogeography of North America, Gondwana, and Intervening Displaced Terranes: Comparisons of Paleomagnetism with Paleoclimatology and Biogeographical Patterns," *Geological Society of America Bulletin*, Vol. 100, No. 3, 311–324 (March 1988).
Vasyutin, V. V., and A. A. Tishchenko: "Space Colorists," *Sci. Amer.*, 84 (July 1989).
Wakita, H., et al.: "Radon Anomaly: A Possible Precursor of the 1978 Izu-Oshima-kinkai Earthquake," *Science*, **193**, 882 (1980).
Wallace, R. E., Editor: "The San Andreas Fault System, California," U.S. Geological Survey, Denver, Colorado, 1990.
Wang, C.-O., and W. C. Burnett: "Holocene Mean Uplift Rates Across an Active Plate-Collision Boundary in Taiwan," *Science*, 204 (1990).
Weintraub, B.: "An Earthquake's Toll in Colonial Jamaica," *Nat'l. Geographic* (Geographics Section), (March 1991).
Wilcock, W. S. D., et al.: "The Seismic Attenuation Structure of a Fast-Spreading Mid-Ocean Ridge," *Science*, 1470 (November 27, 1992).
Wilson, M. E., and S. H. Wood: "Tectonic Tilt Rates Derived from Lakelevel Measurements, Salton Sea, California," *Science*, **207**, 183 (1980).
Wood, B. J., Bryndzia, L. T., and K. E. Johnson: "Mantle Oxidation State and Its Relationship to Tectonic Environment," *Science*, 331 (April 20, 1990).
Yeats, R. S.: "Active Faults Related to Folding," *Active Tectonics*. Panel chaired by Robert E. Wallace. National Academy Press, Washington, D.C., 1986.
York, D.: "The Earliest History of the Earth," *Sci. Amer.*, 90 (January 1993).

EARTHWORK. This term denotes work, the object of which is to alter the surface of the earth to serve some useful constructional purpose. In addition to excavation, building of embankments and trimming of slopes, earthwork also includes the clearing and grubbing of rough land, grading, etc. Excavation of rock and loose rock is usually considered earthwork. Among the most common forms of earthwork are preparation of subgrade for railways and highways, buildings of embankments for hydraulic work and construction of open drainage systems. In the preparation of a roadbed, the original surface of the ground is altered to the required degree by cuts and fills. As far as possible, cut should equal fill, so that the material excavated may be hauled a short distance and used to fill depressions in the proposed roadway. Where the amount of cut is insufficient for filling, the deficiency must be made up by hauling from borrow pits. An excess of cut is deposited on spoil banks.

The vertical dimension of the cut or fill at any station is found by leveling. An engineer's level is set up near the station. The height of the instrument, which is the vertical distance between the line of sight of the level and a reference plane called the datum, is obtained by taking a rod reading on a bench mark or other point of known elevation above the datum. After the height of instrument is determined a rod reading is taken on the ground at the station. This is the ground rod. The difference between the height of instrument and the known grade elevation at the station is the grade rod. The cut or fill is the algebraic difference between the grade rod and the ground rod. The point where the side slope of a cut or fill will intersect the ground surface is marked by a stake called a slope stake. Slope stakes are set before construction is begun.

To measure the amount of cut or fill in earthwork, transverse cross-sections of the cut or fill are measured at regular intervals. These sections are then plotted on paper, and the area computed. Sections are taken close enough so that the volume of earthwork between them is considered that of a prism of length equal to the distance between stations, and area equal to the average of the two sections. This is known as the end area method. A more accurate result is obtained by the use of the prismoidal formula, which states that the volume of the earth equals $\frac{1}{6}(a_1 + 4a_m + a_2)D$, where a_1 and a_2 are the end areas, a_m is the mid-section area, D is the length of the section being measured. Earthwork is usually done on a contract basis and paid for on the basis of volume excavated, or volume of fill. A fill of earth usually shrinks 10–20% upon compacting. Rock may be expected to occupy 15–30% greater volume after excavation than before.

If the volume of cut or fill between any two successive stations is given an algebraic sign (+ for cut and − for fill), the algebraic sum of the volumes starting at an initial station may be represented by a continuous curve called a mass diagram. The abscissa for any point on the curve is its horizontal distance (in units of 100 feet) measured from the initial station; the ordinate is the algebraic sum (in cubic yards) of the volumes up to the point. Plus-ordinates (cut) or plotted above the x-axis, and minus-ordinates (fill), below. (See figure.)



(*Top*) Cross section of a highway in cut. (*Bottom*) Cross section of a highway in fill.

EARTHWORM (*Annelida, Oligochaeta*). Terrestrial segmented worms of many species. They burrow in earth containing organic matter on which they live, coming to the surface only in damp cloudy weather and at night. Their activity in loosening and mixing the soil in fields is estimated to be valuable in crop production.

EARWIG. See **Dermaptera.**

EASTERLIES. See **Winds and Air Movement.**

EASTERLY WAVE. See **Atmosphere (Earth).**

EASTERN TENT CATERPILLAR (*Insecta, Lepidoptera*). Also known as the apple-tree tent caterpillar (*Malacosoma americanum*, Fabricius), this insect eats the foliage of trees and disfigures them with its nests. The insect is most widespread in the northeastern United States, although it is found throughout the country east of the Rocky Mountains, and in parts of California. Similar species are found in the Rocky Mountains and farther west.

Wild cherry trees and apple trees are most often attacked. Peach, pear, plum, rose, hawthorn, and various shade and forest trees are occasionally infested. In spring, their unsightly nests or "tents" are conspicuous on susceptible trees. They are more active in some years than others. They may eat nearly all the leaves of a tree, which weakens the tree, but seldom kills it. Once the caterpillars mature in early summer, they cause no further feeding damage.

One generation of the insect develops in a year. The larvae, or caterpillars, hatch in spring from egg masses, about the time the first leaves are opening. The young caterpillars keep together and spin threads of silken web. After feeding for about two days, they begin to weave their tent in a nearby tree crotch, sometimes joining with caterpillars from other egg masses. As the caterpillars grow, they enlarge the tent until it consists of several layers. In good weather, the caterpillars leave the tent several times per day in search of food, stringing silk after them. In bad weather, they remain between the layers of the tent.

When fully grown, about 6 weeks after hatching, the caterpillar is nearly 2 inches (5 centimeters) long and sparsely hairy. It is black, has a white stripe along the middle of its back, and other white and blue markings. When they are mature, the caterpillars spin cocoons on tree bark, fences, brush, etc. The cocoon is about 1 inch (2.5 centimeters) long and white to yellowish white, depending upon age and presence of caterpillars. Inside the cocoon, the larvae transform to pupae, the resting stage. In early summer, reddish brown moths emerge and the females deposit masses of eggs in bands around twigs. The eggs are covered with a foamy secretion, which dries to a firm, brown covering that looks like an enlargement of the twig. An egg mass usually contains about 200 eggs. The larvae develop inside the eggs, but they do not hatch until spring.

These caterpillar larvae are prey for other insects, toads, and birds. The small wasps (chalcid wasps. etc.) develop as parasites in the eggs, larvae, or pupae. Control chemicals applied to the nests and about one foot (0.3 meter) of surrounding area is effective. The spray should be applied before nests are 3 inches (7.5 centimeters) in diameter. The insects are easy to control by hand removal and burning if only a few trees are infested. Wherever possible, wild cherry trees growing near orchards should be removed.

A similar species, the *forest tent caterpillar (Malacosoma disstria)*, and the *fall webworm (Hyphantria cunea)* are sometimes mistaken for the eastern tent caterpillar. The forest species attacks mostly forest trees and seldom fruit trees. The web of the fall webworm can be distinguished from that of the tent caterpillars by the fact that the webworm's nest is made at the tip of a branch instead of at the crotch.

EAST GREENLAND CURRENT. An ocean current flowing south along the east coast of Greenland, carrying water of low salinity and low temperature. The east Greenland current is joined by most of the water of the Irminger current. The greater part of the current continues through Denmark Strait between Iceland and Greenland, but one branch turns to the east and forms a portion of the counterclockwise circulation in the southern part of the Norwegian Sea. Some of the east Greenland current curves to the right around the tip of Greenland, flowing northward into Davis Strait as the west Greenland current.

EBONY TREE. See **Persimmon Trees.**

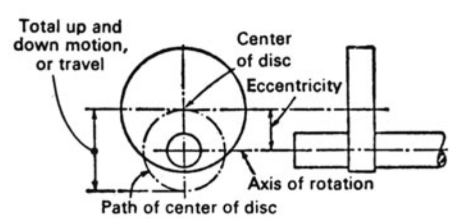
EBULLIOMETER. An instrument, sometimes referred to as an ebullioscope, that measures the property of a substance by noting a deviation from a normal known boiling point. The term applies to apparatus for estimating the percentage of alcohol in a mixture through observation of the boiling point. Beckmann's apparatus for molecular weight determination is an ebullioscope. The ebullioscopic constant is a quantity calculated to represent the molal elevation of the boiling point of a solution by the relationship:

$$K = \frac{RT_0^2}{1000 l_e}$$

in which K is the ebullioscopic constant, R is the gas constant, T_0 is the boiling point of the pure solvent, and l_e is the latent heat of evaporation per gram. The product of the ebullioscopic constant and the molality of the solution gives the actual elevation of the boiling point for the range of values for which this relationship applies. Unfortunately, this range is limited to very dilute solutions, not extending to solutions of unit molality.

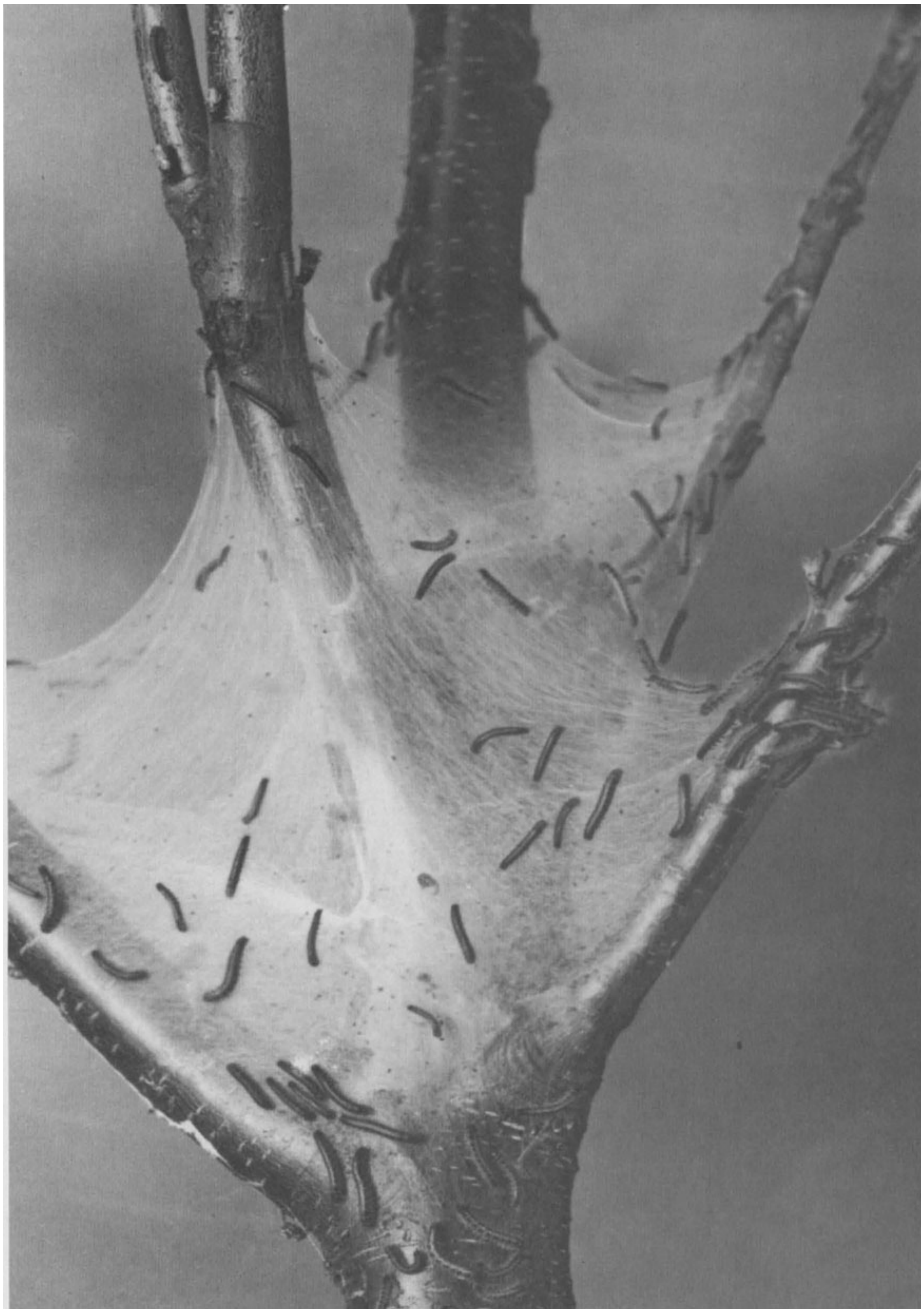
See also **Beckmann Method.**

EBULLITION. See **Boiler; Boiling.**



Simple eccentric.

ECCENTRIC. The eccentric is a machine element employed to convert rotating to reciprocating motion. Its function is similar to that of the crank. The eccentric is used chiefly for short throws, where it would be undesirable to break the shaft, as is necessary in the case of a crank.



Nests of the eastern tent caterpillar. Note the layers of silk.

It consists of a disk mounted on a shaft in such a way that the geometric center of the disk does not coincide with the center of rotation. The distance between the center of rotation and the geometric center of the eccentric is the throw. This corresponds to the crank-arm distance of an equivalent crank. The eccentric must be used in conjunction with an eccentric strap which surrounds the eccentric, and which transmits the reciprocating motion to an eccentric rod rigidly attached. The eccentric is chiefly used to drive auxiliaries such as valve gear, and where reciprocation of small magnitude is needed. The cam and crank may be employed to provide similar motion.

ECHELETTE. The relative intensities of the light in various orders of a diffraction grating depend on the groove shape of the rulings. With only 1,000–2,000 lines per inch (394–787 per centimeter) of ruled surface and proper groove shape, usually rather broad, 75% or more of the reflected light can be thrown into a single order. Such a grating is called an echelette.

ECHELLE GRATING. In studies of high resolution diffraction gratings, it is observed that resolving power at a given wavelength depends only on the ruled width of the grating and the angles of illumination and observation, and not specifically on the number of ruled grooves. The ruling of many inches of a grating with a few tens of thousands of lines per inch is an almost impossible task. An echelle grating has very fine lines ruled much farther apart than is customary. Such a grating has very high resolution, but over only a quite narrow band of wavelengths. Hence it is customary to cross an echelle grating with a second grating (or prism) of lower resolving power, thus producing what is essentially a two-dimensional spectrogram or echellegram.

The resolving power of a diffraction grating depends primarily on the total ruled width and the angles of incidence and refraction. Thus high resolving power should be obtainable with coarse rulings, 100 or so per inch, if the grooves are properly shaped. When the grooves are fixed to reflect light of all colors in a narrow bundle, the resulting echelle grating may have a resolving power equivalent to that of a diffraction grating with 30,000 lines per inch (11,811 per centimeter). It lacks, however, the angular dispersion needed to separate the spectra of consecutive orders; thus this grating is usually crossed with a prism or another grating of lower resolving power.

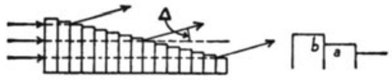
ECHELON. A highly specialized form of diffraction grating, devised by Michelson. It consists of a row of glass plates of exactly equal thickness, packed together to form a miniature stairway of equal risers. The light enters normally to the largest plate at one end (see figure) and emerges at various deviations through the low "risers." It is easily shown that if the thickness of the plates is a, the height of the "risers" b, and the refractive index of the glass n, the equivalent path difference between successive emergent streams for any angle of deviation Δ is $na = a \cos \Delta + b \sin \Delta$; or since Δ is in practice always small, $\cos \Delta = 1$ and $\sin \Delta = \Delta$ (in radians), giving $(n - 1)a + b\Delta$. This must be equal to an integral multiple, N, of the wavelength λ for any spectrum line, the deviation of which is therefore:

$$\Delta = N\frac{\lambda}{b} - (n-1)\frac{a}{b}$$

The smallest value N can have (for $\Delta = 0$) is $(n - 1)a/\lambda$, which, since a is usually several millimeters and $(n - 1)$ is 0.5 or more, is of the order of several thousand. The disposition, viz.,

$$D = \frac{d\Delta}{d\lambda} = \frac{N}{b} - \frac{a}{b}\frac{dn}{d\lambda}$$

is correspondingly large. The echelon is thus especially adapted to the study of the hyperfine structure of spectrum lines. For other instruments of high resolving power, see **Echelette; Echelle Grating;** and **Interferometer.**



Echelon.

ECHIDNA. See **Fossils and Paleontology.**

ECHINODERMATA. A large division of the animal kingdom including the starfishes, sea cucumbers, brittle stars, sea lilies, sea urchins, and basket stars, all marine animals.

This phylum is characterized by the following structures: (1) The adult is almost perfectly radially symmetrical, although the young are bilateral. (2) The wall of the body contains a hard skeleton in most forms, made up of calcareous bodies called ossicles. (3) The coelom is well developed. (4) A water vascular system is present, consisting of a closed series of tubes opening to the exterior at one point on the body and bearing many delicate sacs, the tube feet or tentacles, which protrude at the surface of the body. (5) There is no special excretory system.

The echinoderms are divided into several classes, which fall into two subphyla:

Subphylum *Eleutherozoa*. Without a stalk.
- Class *Asteroidea*. The starfishes.
- Class *Ophiuroidea*. The brittle stars.
- Class *Echinoidea*. The sea urchins, sand dollars, etc.
- Class *Holothuroidea*. The sea cucumbers.

Subphylum *Pelmatozoa*. With a stalk at least when young.
- Class *Crinoidea*. The feather stars, basket stars, and sea lilies.

See also **Invertebrate Paleontology.**

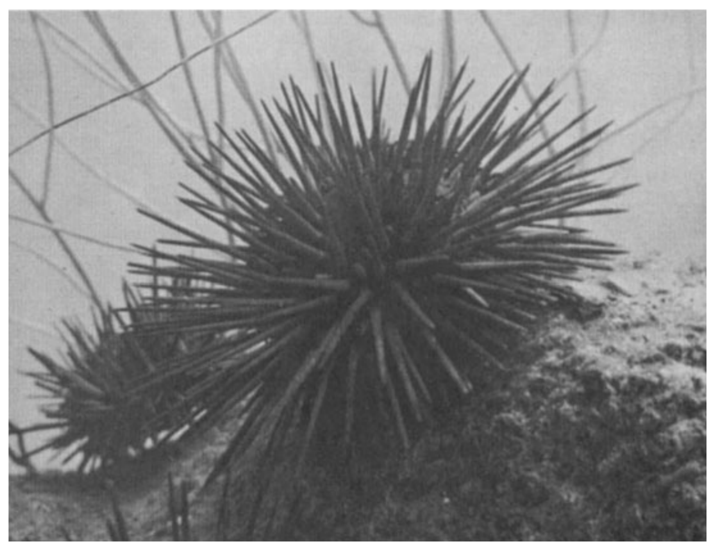
ECHINOIDEA. The sea urchins, keyhole urchins, and sand dollars. A class of the phylum *Echinodermata*.

The members of this class are distinguished by the following characteristics: (1) The body varies from almost globular to thin disks. (2) The tube feet are suckers. (3) The surface bears long spines and pedicellariae. (4) There are no radiating arms. (5) The ossicles are closely associated to form a shell.

Sea urchins live on organic matter of all kinds, including small animals, plant tissues, and waste matter. They are of little economic importance but in some of the Mediterranean countries and to a limited extent in the Orient they are used as food.

The class is divided into the following subclasses:

Subclass I. *Bothriocidaroidea*—extinct.
Subclass II. *Regularia* or *Endocyclica*—periproct encircled by apical system of plates.
Subclass III. *Irregularia* or *Exocyclica*—periproct outside apical system of plates.



Sea urchin (*Echinoidea*). (*A. M. Winchester.*)

ECHO. A wave which has been reflected at one or more points in the transmission medium, or otherwise returned with sufficient magnitude and delay to be perceived in some manner as a wave distinct from that directly transmitted. There are several types of natural echoes, including: 1. The discrete single echo. 2. The discrete multiple echo (a number of successive reflections). 3. The overlapping multiple echo—reverberation. 4. The diffuse echo, due to the scattering of sound by many small objects. 5. The harmonic echo, due to the greater scattering of an overtone than of the fundamental. 6. The musical echo, due to reflection from, or scattering by, a series of objects spaced at uniformly increasing distances from the source. 7. In radar, a general term for the appearance, on a radar indicator, of the radio energy returned from a target. More explicitly, it refers to the energy reflected or scattered back from a target. See also **Sonar.**

ECHOCARDIOGRAPHY. The use of ultrasound in a technique known as *echocardiography* has been available since the late 1960s and has proved of particular value in evaluating such conditions as pericardial effusion, rheumatic disease of the mitral valve, mitral valve prolapse, left atrial tumors, hypertrophic cardiomyopathy, and abnormal dilation of the left ventricle, among other heart problems. See **Heart and Circulatory System (Human).** In heart examinations, the ultrasound beam (narrowly focused) is directed so that it will provide three standard views (left ventricle; mitral valve; aorta and left atrium). The ultrasound reflected from blood-tissue interfaces within the heart and great vessels reveals much concerning the internal anatomy of the heart. The reflected ultrasound radiation is displaced against time, yielding a detailed picture of the motion of the structures of the heart during the cardiac cycle.

In a refinement known as two-sector scanning, a large series of ultrasound beams produces a two-dimensional image of cardiac anatomy.

There appear to be no risks or discomfort associated with echocardiography.

ECHO SOUNDER. A device used to determine the depth of the sea floor beneath sea level. The device emits a high-pitched sound from the ship's hull. Traveling through sea water at a rate of 4,800 feet (1,440 meters) per second, the sound is echoed back from the sea floor to the ship for detection. Since each leg of the round trip is equal, the elapsed time (in seconds) is divided in two and multiplied by 4,800 to give the total depth in feet (or multiplied by 1,440 to give the total depth in meters). Present models can plot a continuous record of depth as the ship travels.

ECHO SUPPRESSOR. When an electric wave on a line encounters a discontinuity or point at which the impedances do not match (see **Impedance Matching**) some of it is reflected. This reflected wave may return to the sending end of the line with sufficient amplitude to be objectionable. This is especially true in telephone service. While an effort is made to prevent reflections, there are cases where energy is fed back along the line and returns to the sender as an echo. In certain systems two lines (4 wires) are employed for transmission in the two directions and in such systems echo suppressors may be used to suppress the returning wave. One rather simple method of achieving this result is to use a relay to short one line when there is a signal on the other. Thus, if party A is talking and sending a voice signal to B, the voice currents on the line from A to B operate a shorting relay across the line from B to A so party A does not receive his own voice as an echo.

ECLIPSE. A term applied to the obscuration of a celestial body due to the interposition of another body or object. There are fundamentally two kinds of eclipse situations, distinguished by whether (1) the eclipsed object is *self-luminous*; or (2) the eclipsed object normally *shines* by *reflected light*.

In the first case, an eclipse occurs when an opaque object passes between the luminous body and the observer. The best known eclipse of this type, of course, is an *eclipse of the sun*, as caused by the moon blocking off light from the sun before the light can reach an observer on earth. Depending upon several factors, such an eclipse may be *total*, where to observers within the *shadow path* of the moon all light is blocked off. Or, such an eclipse may be *partial*, where only part of the light is cut off. Details of a solar eclipse are given a bit later.

In the second case, a body that normally shines by reflected light is eclipsed by putting that body in the shadow cast by an opaque body that intervenes in the direct path from the luminous body and the eclipsed body. The best known eclipse of this type is an *eclipse of the moon*, as caused by the earth blocking off light from the sun before the light can reach the moon. Thus, during the period of an eclipse of the moon, an observer on earth will see the moon pass into and out of darkness. Again, depending upon several factors, such an eclipse may be *total*, where the moon is completely within the shadow of the earth, or *partial*, where only part of the sun's light is cut off. Details of a lunar eclipse are given later.

Eclipses of the first kind also apply to instances where the moon may block off the radiation from a star or reflected light from a planet from reaching earth; and also in the case of eclipsing binary stars. See also **Eclipsing Binary.** Where relatively small objects come between earth and sun, as in the case of a planet blocking off radiation from a star, the term *occultation* is usually used instead of eclipse. Occultation of a planet by the moon also may occur. When an apparently very small object intervenes between sun and earth, as in the case of the planet Mercury, the path of the planet across the face of the sun can be observed. The amount of radiation reaching earth from the sun when this occurs is reduced by such a tiny amount that using the term eclipse is hardly appropriate. For situations of this kind, the term *transit* is commonly used.

Eclipses of the sun and moon have been scientifically important as well as very interesting to lay people and, in fact, to some cultures the eclipses are associated with matters of mystique, superstition, taboos, prophecy, and fear. Mention of eclipses in history is also important to scholars for fixing accurate dates to past events. As an example, reference may be made to an Assyrian table which states; "Insurrection in the city of Assur. In the month of Sivan the sun was eclipsed." This probably also refers to the solar eclipse of June 15, 763 B.C. This is the same eclipse referred to in Amos VII:9: "I will cause the sun to go down at noon, and I will darken the earth in the clear day."

Eclipses, particularly total eclipses of the sun, have been scientifically important in terms of bettering astronomical measurements, the development of solar physics, and proving and disproving various hypotheses and theories, among other scientific objectives. These developments are reviewed briefly a bit later.

Although instruments (see **Sun**) have been developed over the years which essentially duplicate many of the observational advantages of a total eclipse, such as investigating the corona of the sun, large groups of scientists from all over the world continue to travel to areas, sometimes quite inaccessible, to view and carry on special scientific measurements during the period of totality. In recent years, scientists have been joined by growing numbers of lay astronomers who desire the personal experience of witnessing such an event. For example, hundreds of persons, including representatives of the news media, traveled to Kenya's Taita Hills to view the total eclipse of the sun on February 16, 1980.

This occurred again in the case of the 1991 total solar eclipse visible from Hawaii's Mauna Kea.

Geometry of Eclipses

Distances from earth, the relative size or apparent diameters of the participating bodies, and orbital speeds and eccentricities are among the major factors which contribute to positioning the participating bodies at just the right places in the same time frame that result in some kind of eclipse situation. Since all of the foregoing factors relate in a varying manner with time, circumstances for eclipses of the sun and moon occur comparatively infrequently. In only a few years in a given century will circumstances permit up to seven eclipses. In such years, there may be either two of the moon and five of the sun; or three of the moon and four of the sun. Only two years of the 20th century permitted the first of the two foregoing situations—in 1935 and 1982. During this century (1901 through 2000 A.D.), there will have been a total of 375 eclipses, of which 228 will have been solar and 147 lunar eclipses. This is an average of approximately four eclipses per year. Eclipses of the moon, particularly to the layman, may seem more frequent than solar eclipses simply because lunar eclipses are observable over wide areas of earth—in fact, any observer who would be seeing the moon during a given period will witness the moon in eclipse. In contrast, viewing of solar eclipses is limited to a relatively narrow shadow path on earth. Further, the period of totality of a solar eclipse, in the extreme, will not exceed about 7.5 minutes; whereas the moon may be fully darkened up to a period of about 2 hours.

Variations in the sun-earth-moon system not only affect the exact dates on which eclipses will occur, but the type of eclipse (total or partial), the duration of the eclipse, and the locations on earth from which an eclipse can be observed.

With reference to the diagram of Fig. 1 in the entry on **Moon (Earth's),** it will be noted that an eclipse situation can occur only when the moon is at one of its nodes.[1] These nodes occur at those positions when there is a new moon and when there is a full moon. In the idealized diagram of the figure, not allowing for tilt or eccentricity of orbits, these are the points where the moon is most directly between the earth and the sun (new moon); and where the earth is most directly between the moon and the sun (full moon). Total eclipses of the sun occur when there is a new moon; total eclipses of the moon occur when there is a full moon.

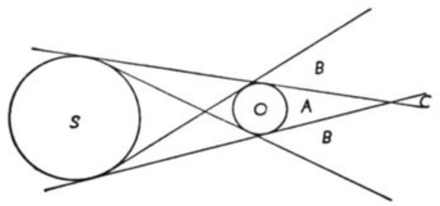


Fig. 1. Formation of shadows for eclipse of sun or moon.

The diagram of Fig. 1 (of the present entry) can be used to illustrate the circumstances of an eclipse of the sun as well as an eclipse of the moon. The opaque body is *O*. The luminous body is *S*.

Eclipses of the Sun

In the case of an eclipse of the sun, Consider *S* to be sun; *O* the moon. The moon will cast a shadow toward earth, known as the *umbra*, and represented by area *A*. The umbra is the darkest central part of a shadow. (When a source of light, not a point source, casts a shadow of an object, the shadow consists of two parts—the umbra, just defined, where the region is completely cut off from the light, and the *penumbra* which is partly illuminated by some part of the light. The penumbra is a region of semi-shadow over which the illumination gradually increases from total darkness to full illumination. From a point within the penumbra, the light source is partially, but not totally occulted by another body.)

A solar eclipse viewed from the region of the umbra will be a *total eclipse*. Viewed from the regions of the penumbra (*B* areas located on either side of the umbra), the occurrence will be seen as a *partial eclipse* of the sun. For an observer within the umbra, the progress of the solar eclipse will appear as diagramed in Fig. 2. For an observer within the penumbra, the disk of the sun will be viewed with a circular segment cut out, thus with the light reduced, but not completely cut off. The viewing path of the solar eclipse of 1979 is shown in Fig. 3.

In an *annular eclipse* of the sun, the moon obscures the central part of the disk of the sun, but leaves a thin ring of light showing round the circumference. With reference to Fig. 1, this is the situation as viewed from *C*. An annular eclipse occurs when the moon is directly between earth and sun, but when the moon is near its farthest point from earth and where the umbra of the shadow does not quite reach the earth. From a knowledge of the diameters of *S* and *O* and the distance between the two objects, the dimensions of the various parts of the shadow may be calculated. The apparent diameters of sun and moon must be nearly the same to assure a total eclipse. This is not true for annular eclipses.

Because the orbit of the earth-moon system about the sun is eccentric, the length of the umbra cone *A* varies between the following approximate distances: At *aphelion* (point where earth-moon system is farthest from the sun), the length of the embra cone is 236,000 miles (377,800 kilometers); and at *perihelion* (point where earth-moon system is closest to the sun), the length of the umbra cone is 228,000 miles (365,000 kilometers). Also, owing to the eccentricity of the moon's orbit, the distance of the moon from the surface of the earth varies between 221,463 miles (356,334 kilometers) at *perigee* (the point where the moon is nearest to the earth) and 252,710 miles

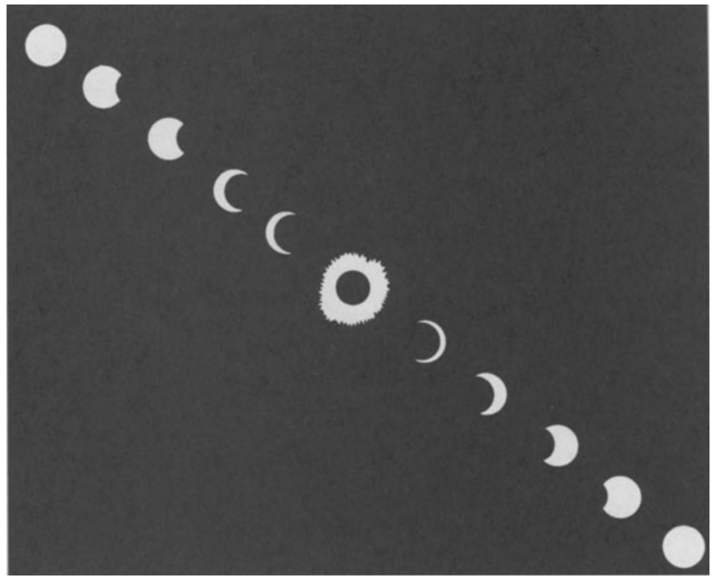


Fig. 2. Progress of a total eclipse of the sun at intervals of approximately 10 minutes.

(406,610 kilometers) at *apogee* (the point where the moon is farthest from the earth). Examination of these numbers indicates that with the earth at aphelion and the moon at perigee, the surface of the earth is about 18,200 miles (29,284 kilometers) inside of the apex of the umbra. Under these conditions, the most favorable for an eclipse of the sun, the shadow of the moon on the earth will be a spot about 170 miles (274 kilometers) in diameter and, within this area, a total eclipse of the sun may be observed. Surrounding the spot of totality, there will be a region of about 3000 miles (4827 kilometers) radius, within which the sun will be partially eclipsed. With the earth at perihelion and the moon at apogee, the surface of the earth will be 19,500 miles (31,376 kilometers) beyond the apex of the umbra cone, a condition permitting the observation of an annular eclipse, but where totality is not possible.

As the moon revolves about the earth in its orbit, the shadow sweeps along the plane of the moon's orbit with a velocity of about 2100 miles (3379 kilometers) per hour to the eastward. The earth is rotating at such a rate that a point on the equator is moving to the east with a velocity of about 1040 miles (1673 kilometers) per hour. Accordingly, under the most favorable conditions for a solar eclipse (moon at perigee: earth at aphelion; eclipse occurring at noon; observer on equator), the shadow will pass the observer with a speed of $2100 - 1040 = 1060$ miles (3379

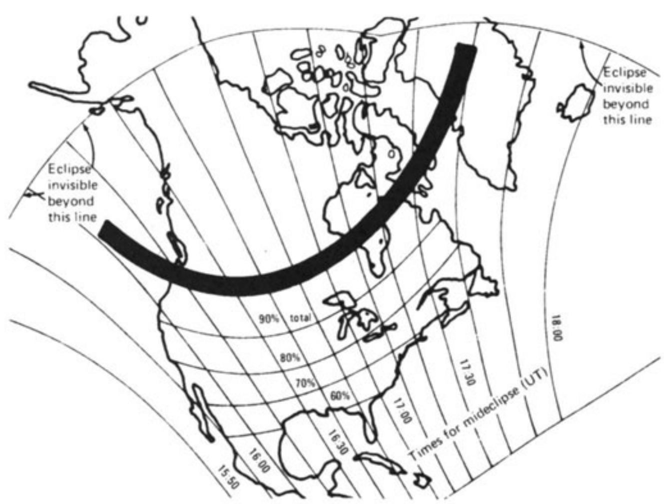


Fig. 3. Viewing path of total solar eclipse of 1979. (*Kitt Peak National Observatory.*)

[1]Nodes are the points at which the orbit of any satellite crosses the plane of the primary's equator or other fundamental reference plane such as the ecliptic. Movement of these crossing points caused by pertubations is termed the *regression of the nodes*.

− 1673 = 1706 kilometers) per hour. The shadow will pass the observer from west to east. The spot will pass the observer in slightly less than 8 minutes. The duration of a partial eclipse, on the other hand, may be several hours.

As stated previously, because the plane of the moon's orbit is inclined to the plane of the ecliptic, an eclipse of either the sun or the moon may occur only when the moon is close to one of the nodes, i.e., close to the plane of the ecliptic, and must also be in conjunction (for an eclipse of the sun), or in opposition (for an eclipse of the moon). Since the earth-moon system revolves about the sun once each year, the line of nodes would pass through the sun twice in each year if the direction of that line were fixed in space. Regression of the moon's nodes causes the line to pass close to the sun three times each year. The period when the line of nodes is close to the sun is known as an *eclipse season*. Two solar eclipses, either total or partial, must occur each year, and five may take place. No lunar eclipse need occur in any year, although three are possible. The minimum number of eclipses in any year is two, both solar.

The sequence of eclipses may be determined by an ancient method known as the *Saros*. The revolution of the moon's nodes is westward at the rate of 19.5° per year. Thus, the sun meets the same node in 346.62 days, the length of an eclipse year. The Babylonians are generally credited for first recognizing the cycle of eclipses. The duration of this cycle is 6585 days (refined to 6585.32 days). This is equivalent to 223 synodic months,[2] and is nearly the same length as 19 eclipse years (6585.78 days). As pointed out by Tver, "After a Saros interval, the sun and moon have returned to nearly the same position relative to each other and to the node, and their distances from earth are nearly the same as before, allowing recurrence of a similar pattern of eclipses." Knowledge of the Saros, as it applied to cycles of lunar eclipses, goes back to about 1000 B.C. The Saros, depending upon the number of intervening leap years, is a period of 18 years, 10.32 days; or 18 years, 11.32 days. The Greeks are accredited with first recognizing the triple Saros cycle of 54 years. This is known as the "exeligmos."

At the end of a Saros cycle, eclipses (both lunar and solar) will recur in the same order and kind (total or partial). However, the essentially identical eclipse will not be observable from the same region on earth—because of the $\frac{1}{3}$ day in excess of the 6585 days. During $\frac{1}{3}$ day, the earth turns 120° and thus the eclipse will be seen about 120° longitude farther west than before. For example, the February 16, 1980 total eclipse of the sun that was observed in parts of Africa (including Kenya), in India, and in southeastern Asia, and of a duration of slightly over 4 minutes, was a repeat of the total eclipse of February 5, 1962, of similar duration and observed in Papua New Guinea. After about 3 Saros cycles (54 years), the eclipses are observable in very near the same longitudes from which they were seen 54 years prior. The number of eclipses in 1 Saros is about 70 (41 solar; 29 lunar). Of the solar eclipses, 10 are total, 14 are partial, and 17 are annular. Because of minor deviations not fully accounted for by the Saros cycle, it is estimated that there are longer cycles. For example, it is estimated that a lunar eclipse may repeat itself about 48 or 49 times, over a series lasting about 865 years. A solar eclipse may have from 68 to 75 returns over a cycle lasting some 1260 years.

As early as 1887, Theodor van Oppolzer prepared a comprehensive table of eclipses for the period through the year 2000. See Table 1. As an indication of the characteristics of a total eclipse of the sun, information pertaining to the October 12, 1977, the February 16, 1980, and the July 11, 1991, eclipses are given in Table 2.

The progress of an eclipse of the sun is designated by a series of "contacts." The first contact comes when the edge of the penumbra *B* (Fig. 1) first touches the sun; second contact when the sun first passes into the umbra *A*; third contact when the umbra leaves the sun; and fourth contact when the last edge of penumbra leaves the sun. Accurate recording of the times of these contacts yields valuable information regarding the complex motions of the moon.

During the progress of a total solar eclipse, about 1 hour before the totality interval, the moon may be seen gradually commencing to cover the sun's disk. About $\frac{1}{4}$ hour before totality, a definite diminution of the light can be noticed and the landscape exhibits hues and tones of color that do not quite seem natural. Most birds and animals seem to sense that something unusual is taking place and become restless. As the crescent of light becomes narrower, images of a crescent shape, known as Baily's beads, may be noticed on the ground, particularly where the diminished sunlight is filtered through the leaves of trees and bushes. These light patches are believed to be caused by the penetration of the last rays of sunlight between the irregularities of the moon's surface. Although not always noticed, within 2 to 3 minutes prior to full disappearance of the sun's disk, shadow bands may appear to be moving over the ground. These bands also are present during the same interval after totality and there have been reports of their presence during totality. Although not fully explained, it is believed that these bands are the result of irregular refraction of light in the earth's atmosphere. Darkness falls suddenly and the white corona seems to flash into position. Almost concurrent with the appearance of the corona an intensely bright region, known as the "diamond-ring" effect may be seen.

The full *corona* of a total solar eclipse is shown in Fig. 4. The corona is the outermost part of the sun's atmosphere which becomes visible during a total eclipse of the sun. The shape of the corona changes periodically in sequence with the sunspot cycles and extends outward some 30 solar radii. Estimated temperature of the ionized gases in the corona range between 1 and 2 million degrees K. Elements such as iron, nickel, and calcium are contained in the gaseous corona. See entry on **Sun** for more detail. Adjacent to the silhouette of the moon during the interval of totality is the *chromosphere*. This is a layer of the sun's atmosphere that lies just above the *photosphere* or visible surface, and below the corona. The lower chromosphere is mainly neutral hydrogen gas at a temperature of about 7500 degrees K. The upper chromosphere contains ionized hydrogen at temperatures up to about 1 million degrees K. Prominences in the form of bright pink spots which extend outward from the chromosphere usually can be observed.

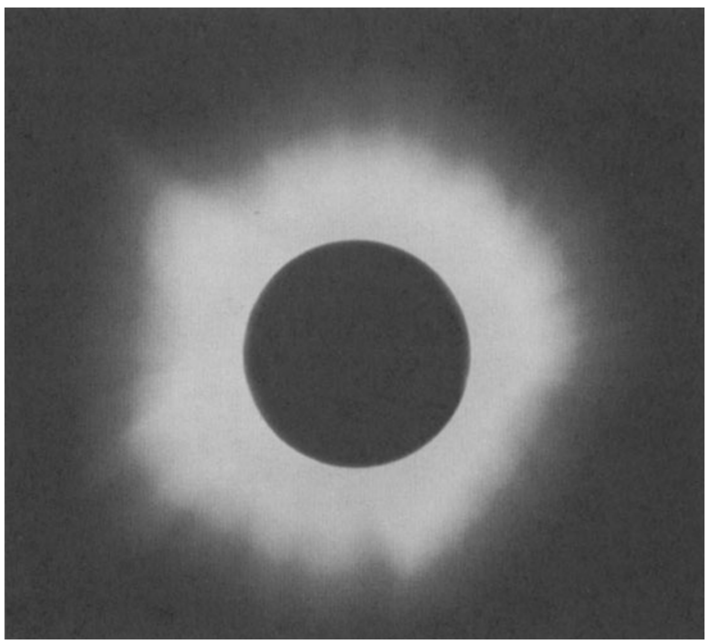


Fig. 4. Total solar eclipse of March 7, 1990, showing full corona. (*Sacramento Peak Laboratory*.)

Considerably dependent upon local climatic conditions, during the period of totality there is usually a noticeable drop in air and ground temperatures. For example, during the February 16, 1980 total eclipse, it was unofficially reported from Ankola, India that the air temperature dropped from 95°F (35°C) to 68°F (20°C) from first contact to totality; while the ground temperature fell from 125°F (51.5°C) to 85°F (about 29°C).

[2]A synodic month is the interval between successive conjunctions of the moon and sun from new moon to new moon again. This interval is a little more than 29.5 days and varies by more than half a day during the year.

TABLE 1. RECORD OF TOTAL ECLIPSES OF THE SUN SINCE 1962 AND FORECAST THROUGH 1999

Year	Month and Day	Location of Shadow Path	Type of Eclipse	Duration (Minutes)
Saros period beginning in 1962				
1962	February 5	Papua New Guinea, Pacific	Total	4
	July 31	Guyana, central Africa	Annular	4
1963	January 25	South Atlantic	Annular	1
	July 20	Japan, Alaska, Canada	Total	1
1965	May 30	South Pacific	Total	5
	November 23	Northern India, southeast Asia, Borneo and the Pacific	Annular	4
1966	May 20	Northern Africa, the former U.S.S.R., and China	Annular	1
	November 12	Pacific, Argentina, South Atlantic	Total	2
1967	November 2	South Atlantic	Total	—
1968	September 22	Arctic, the former U.S.S.R.	Total	1
1969	March 18	Indian Ocean, Pacific	Annular	1
	September 11	Northern Pacific, Peru, Brazil	Annular	2
1970	March 7	Pacific, Mexico, Georgia, north Atlantic	Total	3
	August 31	South Pacific	Annular	7
1972	January 16	Antarctica	Annular	—
	July 10	Siberia, Alaska, northern Canada	Total	3
1973	January 4	South Pacific, South Atlantic	Annular	8
	June 30	Guyana, Atlantic, northern Africa	Total	7
	December 24	Pacific, Brazil, Atlantic, Sahara	Annular	12
1974	June 20	Indian Ocean, southwestern Australia	Total	5
1976	April 29	Algeria, Turkey, southern former U.S.S.R., Tibet	Annular	7
	October 23	Tanzania, Indian Ocean, southern Australia	Total	5
1977	April 18	Atlantic, South Africa, Indian Ocean	Annular	7
	October 12	Northern Pacific, Venezuela	Total	3
1979	February 26	Northwestern United States, Canada, Greenland	Total	3
	August 22	South Pacific	Annular	7
End of Saros period beginning in 1962. Start of Saros period beginning in 1980				
1980	February 16	Congo, Zaire, Kenya, India, southeast Asia	Total	4
	August 10	Pacific, Brazil, Peru	Annular	3
1981	February 4	South Pacific	Annular	3
	July 31	The former U.S.S.R., northern Pacific	Total	2
1983	June 11	Indian Ocean, Indonesia, Papua New Guinea	Total	5
	December 4	Mid-Atlantic, Zaire, Somalia	Annular	4
1984	May 30	Mexico, southeastern United States, Algeria	Annular	1
	November 22	Papua New Guinea, South Pacific	Total	2
1985	November 12	South Pacific	Total	—
1986	October 3	North Atlantic, Greenland	Total	1
1987	March 29	Southern Argentina, South Atlantic, Zaire, Somalia	Annular	—
1988	March 18	Indonesia, Philippines, North Pacific	Total	4
	September 11	Indian Ocean	Annular	7
1990	January 26	Indian Ocean	Annular	—
	July 12	Finland, Siberia, North Pacific	Total	3
1991	January 15	South Pacific	Annular	9
	July 11	Hawaii, Central America, Brazil	Total	7
1992	January 4	Mid-Pacific, California	Annular	12
	June 30	South Atlantic	Total	5
1994	May 10	United States, North Atlantic, Morocco	Annular	7
	November 3	Pacific, middle South America, South Atlantic	Total	4
1995	April 29	Pacific, Peru, Brazil	Annular	7
	October 24	Iran, India, southeast Asia, Pacific	Total	3
1997	March 9	Mongolia, Siberia, Arctic	Total	3
1998	February 26	Pacific, Colombia, North Atlantic	Total	4
1998	August 22	Indonesia, Borneo, Papua New Guinea, South Pacific	Annular	3
1999	February 16	Indian Ocean, northern Australia	Annular	1
	August 11	North Atlantic, central Europe, India	Total	2

TABLE 2. REPRESENTATIVE CIRCUMSTANCES OF RECENT TOTAL ECLIPSES OF THE SUN

1977 (October 12)—Maximum Duration of Totality = 2 min, 27.4 sec

Partial phases visible from all 50 of the United States, western and southeastern Canada, Mexico, and Central America. Path of totality commenced in the Pacific Ocean, ran north and parallel to the Hawaiian Island chain, entered South America in the Darien region of Colombia, and ended in Venezuela.

1989 (February 16)—Maximum Duration of Totality = 4 min, 12.4 sec

Partial phases visible from all parts of Africa except extreme northern part, all of the Arabian Peninsula, all of India, much of southeastern Asia, and China. Path of totality passed from the eastern Atlantic to mouth of the Zaire (Congo) River into Tanzania, passed into the Indian Ocean at Malindi, Kenya, touched land again in Ankola, India, and entered the Bay of Bengal south of Calcutta and ended in China near north latitude 26.5° and east longitude 108.5°.

1991 (July 11)—Maximum Duration of Totality = 6 min, 51 sec

Partial phases visible from western Canada and the United States, Mexico, northern and southern South America as far south as Chile, and wide stretches of the Pacific Ocean. Path of totality commenced in the Hawaiian Islands north-easterly, then passed southeasterly to Mexico City, Colombia, and central Brazil.

This eclipse will be remembered for two major factors: (1) It was, by far, subjected to the most intensive scientific investigation of any solar eclipse to date, and (2) the time of totality was exceptionally long (only 40 sec short of the theoretical maximum). It was the last eclipse to occur with such a long span of totality until the year 2,132.

The path of the total eclipse (umbra) passed directly over Mauna Kea in Hawaii, where telescopes and other solar observing instrumentation is located. Included was the Canada-France-Hawaii (CFH) telescope and, as mentioned by Jay Pasachoff, "the largest ever pointed at the sun." Other instrumentation was brought into play. For example, some 10,000 images of the sun's corona in infrared (IR) wavelengths were made by an imaging system that consisted of an IR camera featuring a 128×128 indium antimode focal plane array, video acquisition and processing electronics, a real-time digital recorder, and image generation and enhancement capabilities. Although the brightness of the corona during the eclipse may have masked faint targets, such as rings, the information gained will assist in updating the multispectral model of the sun's outer surface.

Although, at one time, total solar eclipses were essentially the only way an astronomer could study the outer edges (corona) of the sun, there since have been many advances achieved by way of satellite-based observations. Notably, the Japanese Yohkoh satellite has captured much new information from x-ray imaging of the solar corona and surface. See **Satellites (Scientific and Reconaissance).**

The beams of darknesss that swept across the earth's surface resulting from the eclipse were observed from space for the first time by the GOES weather satellite.

Total Solar Eclipse Research. Considerable attention of a scientific nature to total eclipses of the sun is generally regarded as having commenced with the eclipse of July 8, 1842. The path of that eclipse passed from Spain through France, Italy, Austria, Russia, and central Asia. Along the path were located the leading astronomers of Europe. For the first time, large red solar prominences were reliably described. The extent of the corona was measured. Improvements in the examination of the corona and prominences continued during the total eclipse of July 28, 1851. A successful daguerreotype photo was taken in Königsberg, using a telescope with a 2.4-inch (6-centimeter) aperture. Exposure was 1 minute, 24 seconds. The corona and several prominences were well pictured. Scientists were better prepared when the path of the eclipse of July 18, 1860 crossed through Spain, the Mediterranean, and north-eastern Africa. Photographic and optical techniques had substantially improved during the 9-year interval between these eclipses and the astronomers also had prior knowledge and experience as a guide to the best exposure time to use.

Spectroscopic techniques were first used in connection with the total eclipse of August 18, 1868. The period of total of 5.5 minutes of this eclipse was relatively long. The bright prominences clearly produced several bright lines, this indicating the gaseous nature of the prominences. Identification was made of the hydrogen lines C and F and of a yellow line, later found to be due to helium and designated the D line. Immediately after the 1868 eclipse, it was erroneously identified as a sodium line. The *flash spectrum* is described under **Sun (The).**

The interest of Janssen was spurred by the 1868 eclipse and he soon found that the bright lines of the prominences could be observed on any clear day without waiting for another eclipse. He soon noted the marked and rapid changes that occur in the prominences. Lockyer came up with the same concept at about the same time, and the achievements of both Janssen and Lockyer were reported at the same meeting of the French Academy.

Observations made of the eclipse of August 7, 1869 showed a bright green line across the continuous spectrum of the corona. For awhile, a new element (*coronium*) was proposed as the source of the line, but spectroscopic analyses of the total eclipses of 1896 and 1898 disproved that.

The possible relationship between sun spots and the corona was first investigated during the total eclipse of July 29, 1878, the path of which went diagonally across North America. Langley observed this eclipse from Pike's Peak at an elevation of 14,100 feet (4298 meters). Further evidence of a relationship was noted during the total eclipse-of May 17, 1882.

Coupled with a growing sophistication of instrumentation during the later 1800s and up to the present time were discussions of an increasingly complex nature, based upon knowledge gained from and inspired by eclipse studies. Some of these factors are further discussed in entry on **Sun.**

The availability of greatly improved telescopes and solar instrumentation data gathering techniques, coupled with the long duration of totality of the July 11, 1991, total eclipse, without doubt makes that eclipse the most thoroughly researched of all to date. See Table 2. One of numerous new techniques was application of infrared imaging. See Fig. 5. With the employment today of space satellite observations of the sun, there is less dependence on total solar eclipses for numerous studies.

Precautions in Viewing a Solar Eclipse. Kitt Peak National Observatory astronomers caution skywatchers who plan to observe an eclipse of the sun that *partial and permanent eye damage can be caused by looking at the sun, even for only an instant, without adequate protection*. They further observe, "Sunglasses, smoked glass, welder's goggles, photographic neutral-density filters or color film are *not* safe to use, even in combination. They are not dark enough for protection. A solar eclipse is most safely observed by not looking at the sun at all, but instead by watching its image projected on a piece of paper or board.

"To make your own projector, begin with two white sheets of poster board, or heavy paper. Punch a small, round hole in the center of one sheet. This will serve as your 'lens.' Aim the lens at the sun, holding the second board in your other hand. By moving the two boards, you will find that you can project a sharp image of the sun through the lens onto the second board. Information on how to photograph a solar eclipse can be obtained from the Public Information Office, Kitt Peak National Observatory, P. O. Box 26732, Tucson, Arizona 85726."

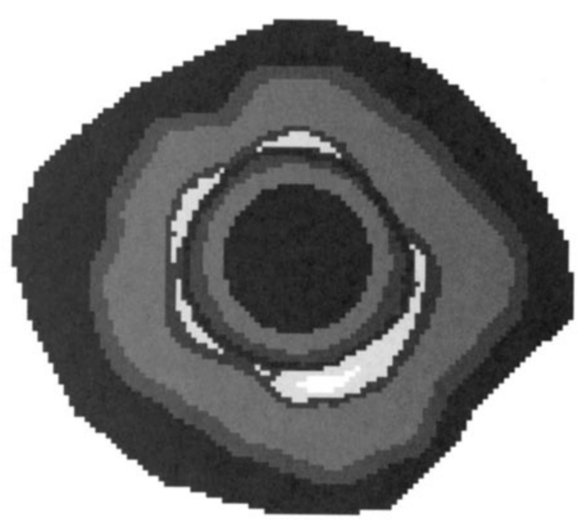


Fig. 5. Facsimile of digitized infrared (IF) image of sun during total eclipse over Mauna Kea, Hawaii, on July 11, 1991.

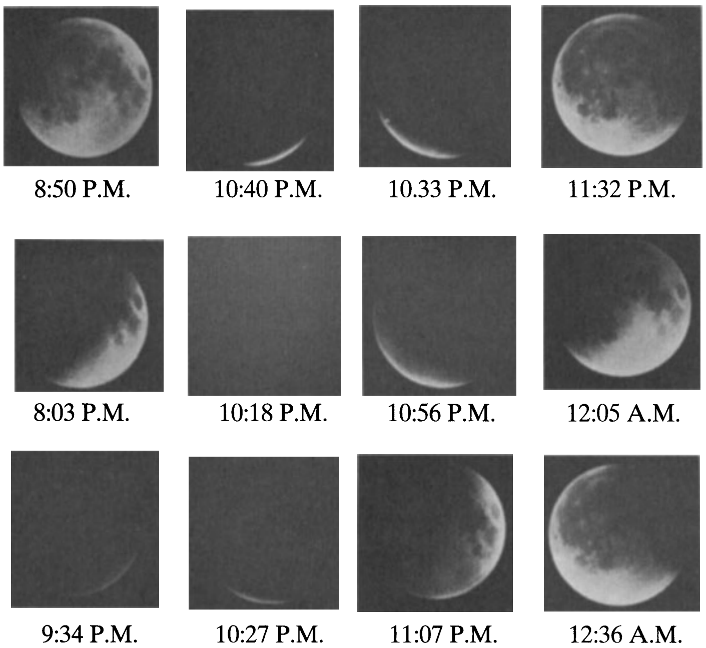


Fig. 6. A lunar eclipse sequence. (*Griffith Observatory.*)

Eclipses of the Moon

Referring back to Fig. 1, for an eclipse of the moon, consider S as the sun and O as the earth. From the relative dimensions and distances, we find that even under the most unfavorable conditions, i.e., with the earth at perihelion and the moon at apogee, the shadow of the umbra cone of the earth will extend well out beyond the distance of the moon. Hence, there are no annular eclipses of the moon. However, partial eclipses, i.e., eclipses that occur when the moon passes through the earth's shadow far enough off the central line to pass outside the umbra, are quite common. These are known as *penumbral* eclipses. The progress of a total lunar eclipse is shown in Fig. 6. Although not visible in Fig. 6, when the moon is in darkness, it is not completely dark during eclipse, but is illuminated by a dull-orange light refracted into the umbra by the atmosphere of the earth. For possibly 30 minutes before the moon reaches the earth's shadow, the darkening of the eastern side will proceed quite gradually. Although the edge of the earth's shadow appears sharp to the naked eye, it appears fuzzy with a field glass or telescope. An eclipse of the moon is visible over an entire hemisphere (sometimes a little more) of the earth and thus for persons in any given locality, a lunar eclipse is more common than a solar eclipse, since even though the latter occur more frequently, they are only visible in much more limited areas. On occasions when the sunlight must pass through clouds, the moon's disk may become completely invisible during the period of totality. Total lunar eclipses in the recent past and up to the year 2000 are listed in Table 3.

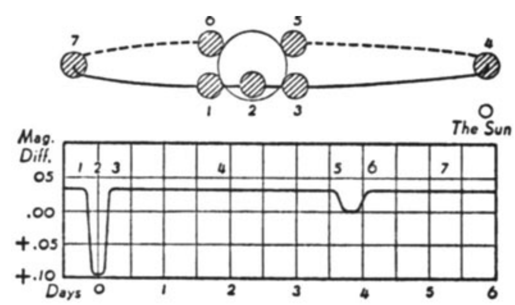
For references, see entry on **Astronomy;** and **Moon (The).**

TABLE 3. LUNAR ECLIPSES THROUGH YEAR 2000

Year	Month and Day	Location Where Moon Is at Zenith	Duration (Minutes) Complete	Duration (Minutes) Total Phase
1982	January 9	Pakistan	214	84
	July 6	Easter Island	224	102
	December 30	Hawaiian Islands	210	66
1983	June 25	Pitcairn Island	130	—
1985	May 4	Mauritius	212	70
	October 28	Bay of Bengal	204	42
1986	April 24	New Hebrides	210	68
	October 17	Arabian Sea	212	74
1987	October 7	Venezuela	22	—
1988	August 27	Samoa	122	—
1989	February 20	Philippines	212	76
	August 17	Central Brazil	220	98
1990	February 9	Southern India	204	46
	August 6	Northeastern Australia	174	—
1991	December 21	Hawaiian Islands	70	—
1992	June 15	Northern Chile	174	—
	December 9	Southern Algeria	212	74
1993	June 4	New Caledonia	220	98
	November 29	Mexico City	206	50
1994	May 25	Southern Brazil	116	—
1995	April 15	Fiji Islands	78	—
1996	April 4	Gulf of Guinea	216	84
	September 27	French Guiana	212	72
1997	March 24	Northwestern Brazil	194	—
	September 16	Maldives	210	66
1999	July 28	Samoa	142	—
2000	January 21	Puerto Rico	214	84
	July 16	Northeastern Australia	224	102

ECLIPSING BINARY. When the orbit plane of a binary star lies so nearly in the line of sight of the observer that the components undergo mutual eclipses, the object is known as an eclipsing binary. If the binary is also a spectroscopic binary, and if the parallax of the system is known, we have one of the most valuable specimens for stellar analysis. Eclipsing binaries are variable stars, not because the light of the individual components vary, but because of the eclipses. The most notable of the eclipsing binaries is the star Algol (β Persei).

The light curve of an eclipsing binary is characterized by periods of practically constant light with periodic drops in intensity. In the Fig. 1 we have a characteristic light curve of such an object. Here, the eclipse of the larger and brighter primary star by the secondary star is annular, and the eclipse of the secondary by the primary is total.



Apparent relative orbit and light curve of the eclipsing binary 1H Cassiopeiae. (*Curve and orbit determined by Joel Stebbins from observations with the photoelectric photometer at the University of Illinois.*)

The orbit of an eclipsing binary may be determined from a study of the light curve. In addition to the seven elements of the orbit, it is also possible to determine the relative sizes of the individual stars in terms of the radius of the orbit. In the determination of the orbit of a spectroscopic binary, it is impossible to determine the semimajor axis, a, and the inclination of the orbit plane, i, independently; but a quantity ($a \sin i$) expressed directly in linear units (i.e., miles or kilometers) may be determined. If a star is both an eclipsing and spectroscopic binary, one can determine all seven elements of the orbit, including a and i, in angular units from the light curve, and the quantity ($a \sin i$) in linear units from the spectroscopic data. Hence, one can determine the radius of the orbit in linear units and then get the sizes of the individual stars in linear units. Since the relative masses of the two stars can be obtained from the period, and since the relative sizes are determinable from the combination of the photometric and spectroscopic orbits, then the densities of the individual stars can be found from relative masses and sizes.

See also **Binary Stars; Spectroscopic Binaries;** and **Visual Binaries.**

S. N. S.

ECLIPTIC. The great circle cut out on the celestial sphere by the plane containing the orbit of the earth. The ecliptic is the fundamental plane for the system of spherical coordinates in which celestial latitude and longitude are measured. The ecliptic is also the reference plane to which the planes of the orbits of all the members of the solar system are referred.

The plane of the ecliptic is inclined to the plane of the equator by an angle of approximately 23°27′, known as the obliquity of the ecliptic. The two planes intersect in a line known as the line of nodes. The points where this line of nodes intersects the celestial sphere are known as the equinoxes. The apparent motion of the sun in the ecliptic about the earth, due to the actual motion of the earth in its orbit, causes the sun to pass through each one of the equinoxes once each year. The point where the sun crosses the equator from south to north is known as the vernal equinox, and the opposite extremity of the line of nodes is the autumnal equinox. Due to precession, the direction of the line of nodes is continually changing relative to the stars. At present, the vernal equinox is in the constellation of Pisces. It is continually moving along the ecliptic, in a direction contrary to the annual motion of the sun, at such a rate that it will complete one revolution of the ecliptic in approximately 26,000 years. See also **Celestial Sphere and Astronomical Triangle.**

ECLOGITE. This is a coarse, granular rock composed chiefly of garnet and pyroxene with subordinate amounts of various minerals such as rutile, magnetite, and apatite. Hornblende sometimes is present, replacing the pyroxene, often to the extent that a garnet amphibolite is produced. The origin of eclogites is obscure, they may result in part from the deep-seated metamorphism of gabbronic rocks, but some may have resulted from crystallization of a primary basic magma under conditions of great pressure. They may represent segregations in a highly basic magma analogous to segregations of basic minerals in granites and other common igneous rocks. Seemingly confirmatory evidence of this idea is found in the chunks of ecologite-like material found in kimberlite in the Republic of South Africa.

ECOLOGY. A branch of biology that deals with the relations of organisms and their environment, including their relations with other organisms. It is an interdisciplinary field, cutting across the life and geophysical sciences. Ecological investigations look into two directions: (1) The nature of environments and the demands which these environments make upon the organisms that inhabit them; and (2) the characteristics of organisms (plant or animal), species, and groups that permit or promote their tolerance of specific environmental conditions. In recent years, particular emphasis has been placed on the studies of groups rather than single species and this has given rise to the term *ecosystem*. Odum defines an ecosystem as "any entity or natural unit that includes living and nonliving parts interacting to produce a stable system in which the exchange of materials between the living and nonliving parts follows circular paths."

Environmental or Habitat Approach. Characteristics of an environment fall into three major categories: (1) *physical*; (2) *chemical*; and (3) *biotic*. In connection with any of these factors, if the presence or absence of a given condition is necessary for sustenance of whatever group, species, etc. that is being considered, such condition is referred to as a *limiting factor*. Physical factors include light, temperature, wind, fire, soil texture, etc. Chemical factors include the composition of the water (pH, dissolved gases and solids, etc.); the composition of the air (pollutants, water vapor, etc.); the composition of the soil (alkalinity, acidity, presence of various elements and compounds); etc. Biotic factors are related to food supply and the presence and behavior of neighboring organisms (predators, parasites, etc.).

Environments can be classified into four major types: (1) *freshwater*; (2) *marine*; (3) *terrestrial*; and (4) *symbiotic*. There are several subdivisions of each.

Freshwater environments are of two principal types: (1) *standing-water* and (2) *running-water habitats*. In the first category, there are ponds, lakes, and bogs; in the latter category, streams, rivers, and springs. Limiting factors of significance include temperature, water clarity, concentrations of oxygen and various salts, and evaporation rate in the case of standing waters.

Marine environments, although similar in many respects to freshwater environments, pose special environmental conditions, including depth, salinity, general presence of greater stability, and less, if any, light at great depths. Subdivisions of marine environments include: (1) the *neritic zone*, the relatively shallow waters of the continental shelf; (2) *oceanic region*, the deep waters beyond the continental shelves, which, in turn, is further subdivided into: (a) the *euphotic* zone, the upper portion of the oceanic region where photosynthesis occurs; (b) *bathyal zones*, depths below the euphotic zone but still located on continental slopes; and (c) *abyssal zones*, depths below the euphotic zone in all other parts of the oceans.

Terrestrial environments are divided into nine or more types, called *biomes*. Examples include deserts, tundras, grasslands, savannas, deciduous forests, coniferous forests, tropical forests, and woodlands. These are described under **Biome.** The ocean is also sometimes referred to as a biome.

The environments described thus far have related essentially to nonorganic factors. In the case of the symbiotic environment, the factors of concern are other organisms. There are numerous examples where two or more species may inhabit a given area of close proximity in the absence of predatoriness. Usually, each species present will contribute in some fashion to the well being of others, although what might be called "stand-off" conditions also are found. See also **Symbiosis.**

Communities Approach. Where the interrelationships of a single individual or species with its environment is studied, the term *autecology* may be applied. In contrast, if entire populations are studied as units, the term *synecology* is used. The latter approach is most popular today. The term *population* is used to describe a group of individuals composed of members of a single species or of several closely associated species that occupy a definite environmental area. Where all of the populations that occupy a given geographic area are studied, the term *biotic community* is used.

The portion of the earth on or in which life exists is termed the *biosphere*. This is a shallow surface region, including the oceans and part of the atmosphere. An ecosystem, as previously defined, can be small and simple or large and complex and, in fact, it is not inaccurate to refer to the entire biosphere as an ecosystem. If life were found in other areas of the cosmos, for example, then it is likely that the first ecosystems to be compared would be those of a scale of the biosphere.

Habitat may be defined as that particular environment in which a population lives. A catfish that likes slow-moving streams and lakes is described as having this particular habitat. Obviously, the variation among habitats considering the tens of thousands of life forms is tremendous, although generalizations can be made. *Niche* is a term used to describe that role played by a population within its community and ecosystem. Numerous factors determine niche—eating habits, predatoriness, etc.

The first branch of ecology to develop beyond the stage of life-history study was the description of vegetation. Its basic method is to study the detailed distribution of vascular plants in terms of communities of various types, the pattern and complexity of which depend largely upon the climate and soil.

The major theoretical contribution of this school is the idea of *seral succession* toward a stable climax. According to this idea, if a new environment is created for terrestrial plants, or an old one drastically changed, the vegetation that first develops on it does not remain unchanged for eternity, but rather alters the environment so that it becomes more suitable for some new and different kind of vegetation and so on. Ultimately, a *climax* vegetation develops which is stable under the prevailing climatic conditions and remains until the climate changes to favor something else, or until some new catastrophe changes the environment drastically once again.

Further study of seral succession showed that vegetation patterns were seldom so simple. The climax community is regarded as a useful concept. Even under constant climatic conditions, some sorts of vegetation are not stable, but undergo local cylical changes. At the other extreme, a few kinds of vegetation replace themselves after disturbance without intervening seral stages. It is clear that the sort of climatic constancy that was implied in the climax idea has not prevailed at least since the beginning of the Pleistocene, and that the climax is best considered an ever-changing end-point toward which vegetation develops rather than a state it actually attains.

Populations. Probably the greatest body of coherent ecological theory has been created by students of populations. The study of field populations has uncovered many interesting phenomena, such as the periodic oscillation of arctic small vertebrate populations and the seasonal changes in abundance of planktonic organisms. These observations have stimulated both laboratory experiments and the development of deductive mathematical theory.

If a small number of organisms is provided with a new unexploited environment, the ensuing population growth curve exhibits a roughly sigmoid form, with an initial phase of very rapid, almost logarithmic increase and a later phase in which the rate of growth gradually declines to zero. Such a population history can be described by a curve of the form:

$$dN/dt = rN\frac{(K - N)}{K} \tag{1}$$

where dN/dt is the instantaneous rate of growth; r the intrinsic rate of natural increase in the absence of crowding; N the population size at any time; and K the maximum population size. A formula of this kind is simple to use and easy to comprehend, and although few populations justify the assumptions on which it is based, it has been used not only for curve-fitting and the description of population growth, but as a point of departure for population theory. The principal developments of importance from it have been the prey-predator equations of Volterra:

$$dN/dt = r_1N_1 - \alpha_1N_1N_2 \tag{2}$$

$$dN_2/dt = \alpha_2N_1N_2 - d_2N_2 \tag{3}$$

and the Gause equations of species interaction:

$$dN_1/dt = r_1N_1\frac{(K_1 - N_1 - \beta N_1)}{K_1} \tag{4}$$

$$dN_2/dt = r_2N_2\frac{(K_2 - N_2 - \beta N_1)}{K_2} \tag{5}$$

In these equations, N_2 is the size of one population and N_1 that of another; α_1 expresses the effect of predatoriness on the prey population, and α_2 expresses its effect on the predator. In the absence of prey, the predator should die at the rate of d_2; K_1 and K_2 are the saturation values of two species grown alone; r_1 and r_2 are the intrinsic rates of natural increase of species N_1 and N_2; and α and β express the inhibitory effects of these species on each other. If $\alpha > K_1/K_2$ and $\beta > K_2/K_1$, then either N_1 or N_2 may win out in competition, the result depending upon the initial concentration of the two species. If $\alpha < K_1/K_2$, and $\beta < K_2/K_1$, the species will coexist. If $\alpha < K_1/K_2$ and $\beta > K_2/K_1$, N_1 will be the sole survivor of competition; and if $\alpha > K_1/K_2$ and $\beta < K_2/K_1$, only N_2 will survive.

Work with models of this sort and with experimental laboratory populations that behave more or less in the ways that the models predict has led to concepts about the *ecological niche*, or the way in which an organism fits into the ecological system of which it forms a part. Some authorities hold that Gause's axiom, which states that two species cannot live indefinitely the same way in the same place under constant conditions, is essentially trivial, while others regard it as one of the great generalizations of ecology. Probably its principal value lies in raising the question of how the niches of apparently similar organisms differ, and so compelling ecologists to examine their material very closely. In this, Gause's axiom is similar to the idea of seral succession, which directs attention to the environmental requirements as well as to the ways in which the environment is changed by them. Both ideas stimulate the acquisition of useful information, although they imply an environmental constancy that is seldom found in nature.

Niche theory has led to renewed interest in the taxonomic diversity of natural communities. MacArthur, assuming that there must be some way in which the niches of organisms in an ecological system did not overlap, developed a model of the relative abundance of the individual species in a natural community. According to this model, the expected abundance of the rth rarest species, where there are n species and m individuals and i is the species rank, is given by

$$\frac{m}{n}\sum_{i=1}^{r}\frac{1}{(n - i + 1)} \tag{6}$$

The predictions of the model have been borne out in a number of taxonomically homogeneous cases, but do not seem to hold generally for natural communities.

A different line of approach has been developed by ecologists interested in practical problems of population management, particularly the exploitation of fish populations. The method has been to start with a formula that is essentially a simple equation of continuity:

$$S_2 = S_1 + (A + G) - (M + C) \tag{7}$$

where S_1 and S_2 are the total weight of the population at the beginning and end of the time under consideration; A is recruitment of new individuals to the population; G is growth of the population; M is natural mortality; and C is capture by fishing.

Such an equation is modified as data become available for a specific case and is gradually refined until it is a very powerful predictive tool. The equations become cumbersome in the process, but the mathematics involved is essentially simple, and electronic computation avoids tedium and human errors.

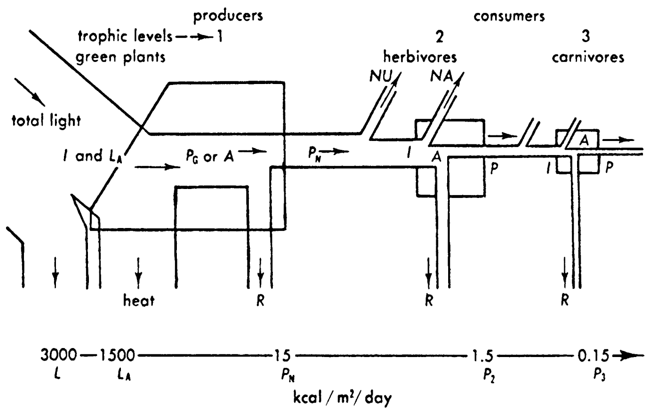
Unfortunately, such models are limited to the populations for which they have been formulated, or for others similar to them in essential respects. There seems little immediate likelihood of the development of simple and accurate general models for the dynamics of natural populations. Many ecologists working with organisms other than fish despair of producing useful models without recourse to stochastic theory. The stochastic approach is appealing, for the processes of population change are certainly not deterministic ones, but such a shift involves a great increase in mathematical complexity. The sophistication necessary to handle stochastic models is rare among ecologists, and it has usually been applied to new methods of constructing a deterministic theory. Cole, for example, used finite difference equations to demonstrate that the reproductive potential of a species depends very greatly upon the timing of reproduction in its life cycle and to point out the critical importance of pieces of information, such as the age at first reproduction, which might otherwise be overlooked in life-history studies. See **Population (Statistics).**

Chemical Aspects. In any community, the interactions of the organisms with each other and with their environment are so complex as to defy complete understanding. One way of asking answerable questions is to restrict attention to the chemical aspects of ecological processes, and to regard organisms as the causes or effects of geochemical processes. This approach has led to an understanding of the quantitative role of organisms in the major chemical cycles, such as the carbon cycle and the nitrogen cycle, and is being strengthened by the use of isotopic methods. The introduction of isotopic tracers into biogeochemical cycles permits the estimation of reservoir sizes and exchange rates and seems likely to lead to new theoretical developments.

Energy Flow. Energy transfer can be considered instead of chemical reactions. See accompanying figure. The principal contribution of the energy flow concept has been to demonstrate that organisms, like most machines, are not efficient converters of energy, so that an organism will incorporate into its body only a small part of the stored chemical energy of its food. This places a severe practical limit on the number of steps that can be maintained in the food chain and has important practical consequences for an expanding human population trying to increase its food supply.

The measurement of energy flow in natural communities presents many difficulties, only some of which are technical in nature. Although the subject is suitable for theoretical development, this has been limited by the lack of methods for dealing with the thermodynamics of open systems. With a better understanding of the thermodynamics of open systems, as being attained by physical chemists, the study of energy flow seems likely to be productive in the future.

Behavior. The experimental analysis of animal behavior patterns has led to a renewed interest among ecologists in the social and psychological aspects of their subject. In particular, studies have focused on terri-



A simplified energy flow diagram. The boxes represent the standing crop of organisms: (1) Producers or autotrophs; (2) primary consumers or herbivores; (3) secondary consumers or carnivores; and the pipes represent the flow of energy through the biotic community. L = total light; L_A = absorbed light; P_G = gross production; P_N = net production; I = energy intake; A = assimilated energy; NA = nonassimilated energy; NU = unused energy (stored or exported); R = respiratory energy loss. The chain of figures along the lower margin of the diagram indicates the order of magnitude expected at each successive transfer, starting with 3,000 kcal of incident light per square meter per day.

toriality, social stress, hierarchies, and other behavioral mechanisms that control population density, on the transmission of traditional information about nesting and feeding sites, and on feeding behavior as it affects an animal's role in the community. Although some behavioral studies have relevance to species diversity, community stability, and population growth rates, much of this work can be carried on profitably outside an ecological context, and behaviorists seem closer to establishing an independent discipline than most other ecologists.

The meaning of the word *ecology* has expanded much during the past few decades. Part of this is due to the fact that ecological matters have broadened and the concerns for ecology have become more intense, thus spawning numerous sub-topics of specialization. Thus, check the Alphabetical Index for such topics as air and water pollution, meteorology, climate, and global change. The latter topic will affect the future ecology of the planet Earth, including the interactions between humans and other life forms. The outline of ecology as given in this particular article can serve as an excellent outline for attacking a myriad of ecological problems.

Additional Reading

Bramwell, A.: "Ecology in the 20th Century," Yale Univ. Press, New Haven, Connecticut, 1989.
Broadus, J. M., and R. V. Vartanov: "The Oceans and Environmental Security," *Oceanus*, 14–19 (Summer 1991).
Ginzburg, L. R., Editor: "Assessing Ecological Risks of Biotechnology," Butterworth-Heinemann, Boston, Massachusetts, 1991.
Jansen, D. H.: "Tropical Ecological and Biocultural Restoration," *Science*, 243 (January 15, 1988).
Ketter, R. B., and M. S. Boyce, Editors: "The Greater Yellowstone Ecosystem," Yale Univ. Press, New Haven, Connecticut, 1991.
Moll, G.: "Designing the Ecological City," *American Forests*, 61–64 (March–April 1989).
Mooney, H. A., and G. Bernardi, Editors: "Introduction of Genetically Modified Organisms into the Environment," Wiley, New York, 1990.
Pimm, S. L.: "The Balance of Nature?" Univ. of Chicago Press, 1992.
Pomeroy, L. R., and J. J. Alberts, Editors: "Concepts of Ecosystem Ecology," Springer-Verlag, New York, 1988.
Price, P. W., et al., Editors: "Plant-Animal Interactions," Wiley, New York, 1991.
Ricklefs, R. E.: "Ecology," 3rd Edition, Freeman, Salt Lake City, Utah, 1990.
Roughgarden, J., May, R. M., and S. A. Levin, Editors: "Perspectives in Ecological Theory," Princeton Univ. Press, Princeton, New Jersey, 1989.
Sherman, K., Alexander, L., and B. Gold, Editors: "Large Marine Ecosystems," American Association for the Advancement of Science, Waldorf, Maryland, 1990.
Vermeij, G. J.: "When Biotas Meet: Understanding Biotic Interchange," *Science*, 1099 (September 6, 1991).

ECONOMIZER. See **Boiler.**

ECOTONE. The often rather blurred and indefinite boundary between two ecological communities. There is usually some tension between two or more communities present at these boundaries. In some instances, the ecotone may represent a fairly wide strip, sometimes referred to as the transition zone. In other instances, the ecotone may be rather sharply marked, as by a stream, edge of a grassland or forest, water hole, etc. See also **Ecology.**

ECOSPHERE. See **Atmosphere (Earth).**

ECTODERM. See **Embryp.**

ECTOPIC. Often as used in medicine, a tumor or growth (or any body tissue) that occurs in an unusual (abnormal) position or in an unnatural form or manner.

ECTOPIC PREGNANCY. See **Pregnancy**

ECTOPROCTA. Members of this group make up the phylum *Bryozoa*. The included species live in colonies of many forms and are characterized by the retractile tentacles and by the anus lying outside of the circlet of tentacles (lophophore).

The phylum is divided into two orders:

Order *Gymnolaemata*. Lophophore circular. Mouth usually closed by a flap called the operculum. Marine.
Order *Phylactolaemata*. Lophophore horseshoe-shaped or oval. Freshwater species.

EDDY. 1. By analogy with a molecule, a "glob" of fluid within the fluid mass that has a certain integrity and life history of its own, the activities of the bulk fluid being the net result of the motion of the eddies. The concept is applied, with varying results, to phenomena ranging from the momentary spasms of the wind to storms and anticyclones.

2. Any circulation drawing its energy from a flow of much larger scale, and brought about by pressure irregularities, as in the lee of a solid obstacle.

EDDY CURRENT. A term generally applied to currents set up in a substance by variation of an applied magnetic field. Eddy currents result in power loss and reduction of magnetic flux. Transformer cores and dynamo frames are laminated to break up the iron structure into thin, insulated layers to reduce the effects of eddy currents. Eddy currents also are used to advantage in induction heating and various damping devices.

Use of separately insulated strands of conductors or bundles of wires in some electrical apparatus accomplishes a reduction in eddy currents just as in the case of using the aforementioned thin, insulated sheets.

EDDY-CURRENT TRANSDUCER. When a nonmagnetic electrically-conductive object is placed in the magnetic field of a coil, the effective inductance of the coil is decreased and its resistance is increased. This results because the field sets up eddy currents in the object that circulate so as to oppose the field that creates them. Moving the object closer to the coil increases the effect. Thus, coil characteristics are a function of spacing between coil and object. This effect is utilized in transducers to sense the position of a specific object or "target."

Various eddy-current transducer systems are available: (1) displacement systems, in which a noncontacting measurement is made of the position of a target, such as a panel being observed for flutter, or a shaft being studied for runout, (2) pressure systems, in which the target is a diaphragm that is displaced by the pressure being measured, (3) accelerometer systems, in which the target is a seismic mass displaced by inertial forces, and (4) differential transformer systems, in which the target is a thin metal sleeve mounted on a ceramic core support that is attached to the mechanism being monitored.

All such systems include electronic circuitry to generate an electrical analog signal related to target position. In some systems, the sensing coil is made part of an impedance bridge circuit so that changes in coil impedance produce an error signal related to target position. In other systems, the coil is part of an oscillator circuit so that the target movements produce frequency changes. Linearity of the output signal may be 1% or less, dependent upon system details. The output voltage may be up to 5 V. The coil excitation frequently may be 1 MHz or higher so that the frequency response of the system can be from 0 to at least 100 kHz. At 1 MHz, the eddy currents are confined near the surface of the target. For example, the "skin depth" in aluminum is 0.0035 inch. Therefore, only a thin target is required and nonconductors can be made into suitable targets by applying, as examples, aluminum foil tape or copper plating.

Operation of eddy-current transducers does not depend upon magnetic properties. Therefore, the devices can avoid problems caused by changes in permeability resulting from temperature changes or stray magnetic fields. Eddy-current transducers are suited to extreme environments and can be fabricated entirely of inorganic and nonmagnetic materials that are highly resistant to the effects of temperature, magnetic fields, and nuclear radiation, including a low cross section of x-rays. The operating principle of the eddy-current transducer is inherently different from the usual magnetic principle, in which coil inductance is increased by an approaching target. These transducers will operate with targets of magnetic materials with reduced, but often with adequate sensitivity.

EDDY VISCOSITY. The turbulent transfer of momentum by eddies, giving rise to an internal fluid friction, in a manner analogous to the action of molecular viscosity in laminar flow (see **Fluid Flow**), but taking place on a much larger scale. If the eddy viscosity is the same at all positions, the equations of motion take the laminar forms but the magnitude of the eddy viscosity must be inferred from other considerations. Values in the atmosphere may be as large as 10^5 cm^2/sec in kinematic units.

EDEMA. Accumulation of excess fluids in the tissues of the body. The condition may rise from several causes. Physiologically the balance of the body fluids between the cells (intracellular fluid), the fluid bathing the cells (extracellular fluid), and the blood plasma is upset. Fluid is drawn from the blood into the tissues when there is a higher osmotic pressure in the tissues than in the blood. This higher pressure may be due to an actual increase (e.g., in salt retention due to impaired kidney function) or it may be a relative increase, as in edema associated with low serum proteins in the blood due to nutritional deficiency. Obstruction to venous flow also results in edema, by the mechanical factor of increased pressure in the capillaries. Capillary damage due to infection, bacterial toxins, or trauma will allow the passage of fluid from the blood into the tissue spaces and produce edema. Exudation of fluid into the extracellular spaces is part of the general process of inflammation and is found at the site of any localized inflammatory reaction or infection.

The common conditions characterized by edema are congestive heart failure, nephritis, varicose veins, cirrhosis, and allergic phenomena such as angioneurotic edema. In congestive failure, the fluid tends to collect in dependent portions, the feet, legs, and over the sacrum. In nephritis these areas plus the loose tissue around the eyes and other easily distensible tissues become edematous. In filariasis, obstruction of lymphatic channels by the parasites results in lymphatic edema, leading to elephantiasis.

See also **Kidney and Urinary Tract.**

EDENTATA (*Mammalia*). This comparatively small order of mammals includes mammals without teeth in the front part of the jaws and with no enamel on the teeth. The feet bear claws. A literal translation of *Edentata* is "without teeth," and thus the term does not seem very appropriate for an order where the majority of animals contained in it do have teeth and, in fact, one of the animals, the giant armadillo, has more teeth than all other mammals with exception of certain species of whales. An approximate classification of *Edentata* includes:

- Anteaters (*Myrmecophagidae*)
  - The Giant Anteater (*Myrmecophaga*)
  - Lesser Anteater (*Tamandua*)
- The Pigmy Anteater (*Cyclopes*)
  - *Note*: The spiny anteaters are of a different order (*Monotremata*); the scaly anteaters are of the order *Pholidota.*
- Sloths (*Bradypodidae*)
  - The Two-fingered Sloth (*Choloepus*)
  - The Three-fingered Sloth (*Bradypus*)
- Armadillos (*Dasypodidae*)
  - Peludos (*Chaetophractus*, . . .)
  - Giant Armadillo (*Priodontes*)
  - The Cabassou (*Cabassous*)
  - Pebas (*Tolypeutes*)

All of the foregoing creatures are considered to be primitive and of very early origin. Although some huge animals of this type once lived, as, for example, ground sloths, they appear to have become extinct during the 16th century. Some of these animals were as large as elephants. At that time, the ground sloths were found in Patagonia and the environs of the Andes. It is also believed that these animals existed in the southwest of North America and some of the islands of the West Indies.

Specializations of the anteaters include a slender elongate snout, a long sticky tongue which aids in gathering a sufficient number of the small prey, and strong claws for tearing open ant nests. Some of the giant anteaters may attain a length of 6 feet (1.8 meters) from head to tail. One pound of ants may be consumed at a single meal. They are gray-black in color, with long hair and a long bushy tail. One young is produced at birth which rides on the mother's back when very young. See Fig. 1.

Fig. 1. Giant anteater, female with young. (*New York Zoological Society.*)

The lesser anteater differs considerably from the giant. The ears are large, the muzzle is short, the tail is opossum-like and, although appearing naked, it is covered with dense, hard fur of various colorations. The tail is used in a prehensile fashion. The animal prefers living in trees where it feeds on ants and termites. They have a reputation of being very strong for their size and quite agile when on the ground. They are considered to be excellent defenders of themselves, usually fighting from a sitting position with their strong arms and sharp claws slashing any predators.

The pigmy anteater is short, about 1 foot (0.3 meter) in length, and inhabits the forests of Trinidad and South America. It is of reddish-brown coloration, has small ears, a short snout, and is seldom seen.

Sloths are animals of Central and South America, highly specialized for arboreal life. Their claws are large hooks by which they suspend themselves from branches, and they are so thoroughly adapted to this inverted position that the hair runs from belly to back, opposite to its direction in animals who maintain the usual position. Sloths eat the foliage of cecropia tress by preference.

The sloths are of two forms: The two-fingered (or sometimes called two-toed) and the three-fingered. Actually, there is a need to differentiate between these two types, but the names are extremely misleading because both types have five toes! The so-called two-fingered sloth is about the size of a large domestic cat, covered with very shaggy and coarse fur along the back. The fingers are ideally shaped, hardly requiring flexing for clinging to vines and tree boughs. The animals prefer leaves and fruits. They are known for their durability and ruggedness and can recover from extremely bad wounds and injuries. When provoked, they also can inflict very bad bites and tears. Generally, however, they are considered rather lazy and mindful of their own business.

The three-fingered sloth or *ai* is a smaller animal, with a dense and hard hair coat. They are of a silver-gray coloration and have a bright yellow or cream-colored face. They have an unusual bright yellow and black sunburst type of marking between their shoulder blades. Although the marking may be coincidental, it is interesting to note that the marking is quite similar to the flower clusters of the wild pawpaw tree on which the animals feed. The young, one at a birth, are tiny. Some experimenters believe that the ais can sense color because they refuse to eat artificially-colored cecropia flowers, or any food that is colored by artificial lighting.

Armadillos are burrowing animals with many bony plates in the skin which form a more or less complete armor when the animal rolls up. Several species occur from Argentina northward through South America, and one, the nine-branded armadillo, is found in the southwestern United States. They range in size from the 5-inch (13 centimeters) pichiciago to the 3-foot (0.9 meter) giant armadillo. As with the other animals previously described in the order of *Edentata*, the mother armadillo carries the baby on her back and in case of trouble rolls around it, assuming the shape of a ball wherein the baby is completely hidden. Some of the particular varieties of armadillo include: the peludo (the hairy armadillo of Argentina, *Tatu pilosa*); the tatoquay (the broad-banded armadillo of South America, *Cabassous unicinctus*); the pichi (the pigmy armadillo, *Chalamyphorus*, of Argentina); the peba (the nine-banded armadillo, *Tatusia novemcinctus*, found from southern Texas and New Mexico to Argentina); and the apar (the three-banded armadillo, *Tolypeutes*, of South America). See Fig. 2.

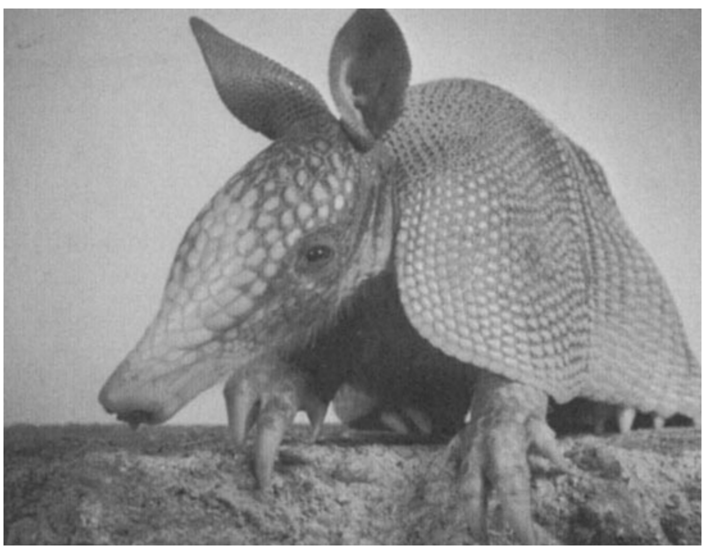


Fig. 2. Armadillo. (*A. M. Winchester.*)

EDGE. Two distinct points (endpoints) and a line segment joining them. Considering the edge as point-closed (including endpoints) is the conventional procedure although open line segments are also used by some authors. Similarly the insistence upon the distinctness of the endpoints of an edge is motivated by the desire to exclude closed "loops" in which the endpoints coincide.

See **Graph (Mathematics); Tree (Mathematics); Vertex.**

EDGE DISLOCATION (Crystal). See **Crystal.**

EDGE EFFECT. In a capacitor comprising two parallel plates, the electric field is normal to the plates except near the edges, where the field lines bulge outward. This edge effect introduces a correction to the capacitance as computed from the parallel field idealization. By giving one plate greater area than the other, and surrounding the smaller plate with an auxiliary guard ring maintained at the potential of the smaller plate, the edge effect is eliminated, and the capacitance between the small plate and the large plate is given by the simple theory.

EDGE SEQUENCE. A subgraph whose edges admit an ordering possessing the following property: Each edge has one vertex in common with the preceding edge and the other vertex in common with the succeeding edge. The words "preceding" and "succeeding" are defined with respect to the ordering imposed on the edges.

EDGE TONES. The tones produced by the splitting of an air-jet by a sharp edge maintained in the jet.

EDGEWORTH SERIES. A form of expansion of a frequency function in terms of derivatives of the normal distribution. For a distribution in standard form (i.e., with zero mean and unit variance), the expansion is

$$f(x) = \left[\exp\left\{ \sum_{j=3}^{\infty} k_j \frac{(-D)^j}{j!} \right\} \right] \alpha(x),$$

where k_j is the jth cumulant; D is the operator d/dx and $\alpha(x)$ is the normal function $e^{-1/2x^2}/\sqrt{(2\pi)}$.

See also **Gram-Charlier Series.**

EDTA (Ethylenediaminetetraacetate). See **Chelates and Chelation.**

EDT CRYSTAL. See **Piezoelectric Effect.**

EEL GRASS (*Zosterna marina*; Najadaceae). Eel grass is a common plant of sandy or muddy ocean shores. The plant has a creeping somewhat fleshy stem which roots freely at the nodes and has short erect branches. Usually it grows in salt water from a foot to over 4 feet (1.2 meters) deep, and is frequently found in tidal pools.

Quantities of eel grass are raked up and dried. It is used for packing, for stuffing for various objects, in the manufacture of certain kinds of wall board and for insulation in walls.

The plants suffer periodically from a certain disease which seriously depletes their numbers. In some regions the natural growth of eel grass has been practically wiped out by this disease.

Freshwater eel grass, *Vallisneria spiralis*, is an entirely different plant, with an interesting method of pollination. It is frequently used in aquaria.

EELS (*Osteichthyes*). Members of the order *Apodes*, eels are elongate slender fishes without pelvic fins and in some species lacking pectoral or median fins.

Eels fall into several classifications: (1) freshwater eels (family *Anguillidae*); (2) parasitic snubnosed eel (family *Simenchelidae*); (3) moray eels (family *Muraenidae*); (4) snake eels (family *Ophichthidae*); (5) snipe eels (family *Nemichthyidae*); (6) deep-sea eels (family *Synaphobranchidae* and family *Serrivomeridae*); (7) conger eels (family *Congridae*); and (8) worm eels (family *Moringuidae*). Also, of the order *Heteromi*, there are deep-sea spiny eels. The latter are not true eels. The so-called electric eel is not an eel. See **Gymnotid Eels (Osteichthyes).**

European Eels. The anguillid eels are found in a number of freshwater areas, such as lakes, ponds, rivers, and streams. See accompanying illustration. To spawn, they return to the sea. In the European eel, well known as an edible fish, females do not spawn until they have reached an age of about 12 years (length of about 5 feet; 1.5 meters). The males migrate for spawning much earlier—when they are from 4 to 8 years old (about 20 inches; 51 centimeters long). During this migration, the eels change color—from their normal yellow to silver and, at this time, they also are quite fat and thus most attractive from a food standpoint. Thus, the European fisheries gear their operations to catching the eels during the downstream migration. By instinct, the eels head for the Sargasso Sea in the Bermuda region. They spawn after having traveled thousands of miles, requiring about a year. After spawning, the adults die. The eggs float to the surface from a spawning depth of some 1,500 feet (450 meters). After hatching, the leptocephalous larvae commence the return trip, requiring in this case some 3 years. Usually arriving in European waters in the spring, the young elvers (about 3 inches; 7.5 centimeters long) seek out fresh water. The designers of the Zuider Zee dike did not contemplate the life cycle of the eel and hence millions of the elvers perished until so-called elver runs were created especially for them.

The mystery of eel migration was studied extensively by Danish biologist Johannes Schmidt who, in 1906, proposed a solution to the puzzle. From his investigations, he learned that the American eel (*Anguilla rostrata*) spawns in about the same area as the European eel (*Anguilla anguilla*). The puzzling problem was the difference in time required for the return trip, one year for the American eel; about three years for the European eel. The American eel can be found northward from Brazil to Greenland and Labrador and has penetrated into the central portions of the United States.

European eel (*Anguilla anguilla*).

There are several species of freshwater eels, but none have been identified in the South Atlantic or eastern Pacific. These latter waters are not sufficiently saline and at the desired temperature at spawning depths.

The blood of the anguillids contains a toxin which affects the nerves and penetrates a wound if exposed during handling and preparation of the eel for food.

Snubnosed Eels. The parasitic snubnosed eel habitates waters at depths between 4,000 and 5,000 feet (1200 and 1500 meters). Maximum length is about 2 feet (0.6 meter). Found in the waters along the American Atlantic coasts, Japan, and the Azores, the *S. parasiticus* obtains its nourishment by drilling into the bodies of larger fishes.

Moray Eels. There are numerous species of moray eels, of which most attain a length up to 5 feet (1.5 meters), although some have been reported as long as 10 feet (3 meters). Although a number of species are poisonous (with a number of deaths recorded as the result of eating), other species are routinely eaten. The several species of moray eels differ considerably in appearance, due mainly to their rostral development and nostril profile. Some appear as dragons and sea monsters. They prefer temperate or tropical waters and dwelling in a rock or reef environment.

Spiny Eels. The deep-sea spiny eels (*Heteromi*) look like eels, but are not related to eels. They are found in very cold waters at considerable depths—in excess of 1000 feet (300 meters) and sometimes reaching a depth of nearly 9000 feet (2700 meters). Some species possess luminous organs along their sides and on their head. They are highly elongated fishes and are seldom seen because of the great depths which they prefer.

Congers. The many genera and species of conger eels (*Congridae*) are distributed in almost all tropical and subtropical seas. The best known of them is the conger eel (*Conger conger*), which reaches a length of about 10 feet (3 meters) and a weight of about 243 pounds (65 kilograms). Distribution is almost worldwide, since it is caught in every ocean except the eastern Pacific. The species prefers rocky coasts, in the crevices of which it hides during the day. It is occasionally caught in brackish water of river mouths. The powerful teeth reflect that this is a predator which is dangerous. Its diet consists of various fishes, crustaceans, and squid. In addition to size, this species is distinguished from the European eel by the longer dorsal fin, which has its base shortly before that of the pectoral fins, and by the scaleless skin.

Not all is understood concerning the spawning of conger eels. Occasionally, females have been captured which have not yet emptied all their contents. Females have also reached maturity in aquariums. After maturity, it is estimated that the highest number of eggs which a female can lay is about 8 million. Pathological changes take place concomitantly with maturation of the gonads, these changes being in the intestinal tract and other organs, including skeleton and teeth. The adults cannot recover from these post-spawning changes and die as a result of them. Eels probably spawn in the open sea at depths of about 8350 feet (2500 meters), but specific spawning grounds have not been found for this group. All other genera and species of this family are considerably smaller than the congers and are almost exclusively deep-sea inhabitants.

Garden Eel. These eels (*Heterocongridae*) are classed with congers by a number of zoologists, while others treat them as an individual family. Garden eels are some 12 to 20 inches (30 to 50 centimeters) long and live in tubes, which extend vertically about 1.5 feet (0.5 meter) deep in loose sand or fine-grained coral sand. It is quite an experience for a diver to encounter such a garden eel "settlement." They often cover 120 square yards (100 square meters) of sandy bottom and the garden eels inhabit the floor at intervals of 8 to 24 inches (20 to 60 centimeters). With a slightly bent fore-end, which protrudes about two-thirds out of the tube, the eels sway to and fro with the head directed against the current, seeking the zooplankton upon which they feed. Such garden eel colonies have only been found at places constantly covered with water, and only where the current is uniform. The garden eels also avoid areas where breakers occur. It is reported that these eels flee from a diver by sliding into the soil when the diver reaches a distance of about 10 feet (3 meters) from them, and only their heads protrude. At a distance of about 3 feet (1 meter), they withdraw completely into their tubes and thus wait for about 5 minutes before shyly looking out again. Most attempts to dig healthy garden eels out of the sand fail, since the animals can dig back into the soil very rapidly. With a poison solution, they can be readily driven out of their tubes. As long as they are uninjured, they swim headfirst with undulating motions over the sandy bottom, in a completely flat position. But after swimming about 3 feet (1 meter) they turn about quickly and rapidly bore tail-first with powerful movements into the sand. Thus, the garden eels are excellent in their maneuvers to avoid capture. Secretion of a mucus substance enables the eels to maneuver in and out of their tubes with ease.

See also **Aquaculture;** and **Fishes.**

EELWORM (*Nematoda*). 1. A group of roundworms of the genus *Heterodera*, parasitic on the roots and underground stems of vascular plants. They cause serious damage to many cultivated crops—potato, sugar beet, beetroot and cereals. Resistant cysts are formed, viable for many years, but the root-knot eelworm causes galls to appear on the roots of tomato, cucumber, and others grown under glass. Entirely satisfactory chemical control has not been achieved as yet, but biological methods using crop rotation have been more successful and economical. 2. The roundworm *Ascaris lumbricoides*, parasitic as adult in the intestine of man and domestic animals, is called an eelworm in some countries.

EFFECTIVE TEMPERATURE (Astrophysics). A measure of the temperature of a star deduced by means of the Stefan-Boltzmann law, from the total energy emitted per unit area. Compare brightness temperature, color temperature. Effective temperature is always less than actual temperature.

EFFICIENCY. The general significance of this term as applied to a device or machine may be expressed as the ratio of output to input of energy or of power. If a dc motor, for example, is operating on 4 amperes at 100 volts (the power input is 400 watts), and if the motor actually delivers only 280 watts of mechanical power, its efficiency at that load is 280 watts ÷ 400 watts, or 70%. In general, the efficiency of a machine varies somewhat with the conditions under which it operates. Usually there is a load for which the efficiency is a maximum. This may be illustrated by a heavy block-and-tackle. For a small load the efficiency would be very low, because of power wasted in bending the ropes; for an excessive load it would again be low, on account of the large friction which would then develop; while for intermediate loads, higher efficiencies would prevail. The concept may be extended to other than purely mechanical systems. Thus the efficiency of an electric lamp may be expressed in candles or lumens of luminous flux (output) per watt of electric power (input); or that of an automobile horn in watts of acoustic power (noise) per watt of electric input. Various types of heat engine exhibit different thermodynamic efficiencies, i.e., the ratio of the work derived in the engine to the heat energy applied to it. See **Thermal Efficiency.** In statistics, the relative efficiency of two consistent estimators t_1, t_2 of a parameter ϕ is defined as the limiting value of the inverse ratio of their variances as the sample size increases. It can be shown that no statistic can have a large-sample variance less than a certain lower bound; estimators whose large-sample variance achieves this lower bound are said to be fully efficient or simply efficient. See also **Machine (Simple).**

EFFLORESCENCE. When a substance evolves moisture upon exposure to the atmosphere, the substance is said to be efflorescent, and the phenomenon is known as efflorescence. At ordinary temperatures, the vapor pressure of water is shown by the accompanying table. If the substance has a higher water vapor pressure than corresponds to that of the atmosphere at the given temperature, water vapor is evolved from the substance until the water vapor pressure of the substance equals the water vapor pressure of the surrounding atmosphere.

Substances that are ordinarily efflorescent are sodium sulfate decahydrate, sodium carbonate decahydrate, magnesium sulfate heptahydrate, and ferrous sulfate heptahydrate. When the saturated solution of a substance in water has a water vapor pressure greater than that of the surrounding atmosphere, evaporation of the water from solution takes place.

See **Deliquescence** for the converse phenomenon.

VAPOR PRESSURE OF WATER

	Water Vapor Pressure in mm Mercury	
Temperature, °C	At Saturation	At 50% Humidity
0	4.6	2.3
10	9.2	4.6
20	17.5	8.8
30	31.8	15.9
40	55.3	27.7

EFFLUENT STREAM. A stream that flows out of another stream or out of a lake. Also a stream whose upper surface is below the surface of the local ground water table. The term *effluent* is frequently used in the processing field to indicate a gaseous or liquid stream that is discharged after some form of treating or processing.

EFFUSION. Effusion is a general term denoting a process of discharge, that is also used specifically to denote the passage of gas under pressure through a small orifice.

EFFUSIVE. The term applied by geologists to molten material (lava) which has been poured out on the surface of the earth from a vent or fissure, as distinguished from ejected volcanic material (ashes and bombs) and injected magmas (plutonic rocks).

EGG. See **Poultry; Protein.**

EIDER DUCK. See **Waterfowl.**

EIGENFUNCTION. If a differential or integral equation possesses solutions satisfying the given boundary conditions for only certain values of a parameter λ, such a value is an eigenvalue (proper or characteristic value) and the corresponding solution is the eigenfunction belonging to that eigenvalue. Thus, given the linear operator P, the solutions u to the equation $Pu = \lambda u$ are eigenfunctions of P belonging to the eigenvalue λ. The totality of eigenvalues of any linear operator constitute the complete set, which may be made orthonormal.

Matrix methods are often useful in discussing this subject. If λ is a scalar parameter, **A** a square matrix of order n, and **E** the unit matrix of the same order, then $\mathbf{K} = [\lambda\mathbf{E} - \mathbf{A}]$ is the characteristic matrix of **A.** The equation $det\, \mathbf{K} = 0 = \lambda^n = a_1\lambda^{n-1} + \cdots + a_n$, where the a_i are functions of the elements of **A,** is the characteristic equation of **A** and its roots are the eigenvalues, characteristic or latent roots. The trace of **A,** the sum of its diagonal elements, is the sum of the eigenvalues. Two matrices related by a similarity transformation have the same eigenvalues and hence the same trace.

An eigenfunction is often regarded as a vector in an abstract n-dimensional space; hence, it is often called an eigenvector.

EIGENVALUE (Proper Value). The concept described by this hybrid word has become extremely important in pure and applied mathematics, and in engineering, physics and chemistry. The German word "*eigen*" means "characteristic," a term already overburdened in mathematical English. The "characteristic values" of a physical system are numbers which describe, for example, the critical frequencies of a suspension bridge or of a rotating shaft, the critical load of a supporting column, or the energy levels of a system in quantum mechanics. In corresponding mathematical language, we shall define the eigenvalues of a square matrix, of the kernel of an integral equation, of a differential equation with boundary conditions, and of an operator or transformation, the last case including all the others.

If $\mathbf{A} = \{a_{ik}\}$ is a square matrix, then the number λ is an eigenvalue (also called a characteristic number, a proper value, a latent root, etc.) of the matrix **A** if the determinant

$$\begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix}$$

is equal to zero. In other words, an eigenvalue of a matrix is a root of the characteristic equation of the matrix.

As a physical example, consider a suspension bridge which is vibrating with n degrees of freedom about a position of stable equilibrium. Its potential energy at any time will be a function, call it $A(q_1, q_2, \ldots, q_n)$, of its n position coordinates $q_1, q_2, \ldots, q_n$, which is approximated by the quadratic form $\Sigma\, a_{ik}q_iq_k$, the constants a_{ik} being given by $a_{ik} = \frac{1}{2}(\partial^2 A/\partial q_i \partial q_k)$ evaluated at the position of equilibrium. Then, the eigenvalues of the matrix $\{a_{ik}\}$ are the squares of the critical frequencies of the bridge, namely the frequencies at which an impressed force will cause dangerous resonance.

From the theory of linear equations it follows that this definition of the eigenvalues of a matrix can be restated as follows. The constant λ is an eigenvalue of the matrix $\{a_{ik}\}$ if the n homogeneous linear equations in the n unknowns $q_1, q_2, \ldots, q_n$

$$\begin{aligned} a_{11}q_1 + a_{12}q_2 + \cdots + a_{1n}q_n &= \lambda q_1 \\ a_{21}q_1 + a_{22}q_2 + \cdots + a_{2n}q_n &= \lambda q_2 \\ \vdots \quad\quad \vdots \quad\quad \cdots \quad\quad \vdots \quad &\quad \vdots \\ a_{n1}q_1 + a_{n2}q_2 + \cdots + q_{nn}q_n &= \lambda q_2 \end{aligned}$$

have a non-trivial solution (i.e., one for which not all the q_i vanish). In other words, regarding the matrix as a linear operator in n-dimensional Euclidean space, the constant λ is an eigenvalue of the operator A which transforms a vector $\mathbf{q} = (q_1, q_2, \ldots, q_n)$ into a vector denoted by $A\mathbf{q}$, if there exists a non-zero vector $\mathbf{q} = (q_1, q_2, \ldots, q_n)$ such that $A\mathbf{q} = \lambda\mathbf{q}$. Such a non-zero vector is called an *eigenvector* (of the operator A) belonging to the eigenvalue λ.

Similarly, let $A(x, y)$ be a continuous function of two variables (the infinite-dimensional analogue of a matrix with rows and columns) defined for $a \leq x, y \leq b$, and let A be the integral operator which transforms a function $\phi(x)$ into a function $f(x)$ according to the formula

$$f(x) = \int\int A(x, y)\phi(y)\,dy,$$

where the integration is taken over the square $a \leq x, y \leq b$ and the functions $\phi(x)$ and $f(x) = A\phi(x)$ are infinite-dimensional analogues (in Hilbert space) of the above vectors $\mathbf{q}$ and $A\mathbf{q}$. Then the constant λ is defined to be an eigenvalue of the integral operator A if the integral equation

$$\int\int A(x, y)\phi(y)\,dy = \lambda\phi(x),$$

which we may also write in the form $A\phi = \lambda\phi$, has a non-trivial solution $\phi(x)$. The solution $\phi(x)$ is called an eigensolution, or eigenfunction belonging to λ.

The type of eigenvalue problem occurring most frequently in practical applications involves a differential operator. For example, the critical frequencies of a vibrating plate (important in the theory of microphones and elsewhere) are the squares of the eigenvalues of the biharmonic operator discussed below.

In general, let A be any differential operator acting on functions ϕ of any number of variables; e.g., for the clamped plate, $\phi = \phi(x, y)$ and A is the biharmonic operator defined by the formula

$$A\phi = \partial^4\phi/\partial x^4 + 2\partial^4\phi/\partial x^2\partial y^2 + \partial^4\phi/\partial y^4.$$

Also, let us consider as admissible only those functions ϕ which satisfy certain boundary conditions; e.g., for the clamped plate, ϕ and its normal derivative must vanish on the boundary. Then the constant λ is an eigenvalue of the operator A if there exists an admissible function ϕ, called an eigenfunction of A, such that $A\phi = \lambda\phi$. Differential eigenvalue problems of this sort are of fundamental importance in the theory of vibrations and buckling and in quantum mechanics.

EIKONAL EQUATION. A fundamental equation of wave motion of the form:

$$|\nabla\psi|^2 = n^2(x, y, z)$$

where n is the refractive index of the medium for the waves, ∇ is the vector differential operator, and $\psi(x, y, z)$ is a function called the eikonal, which defines the wave fronts.

EINSTEIN DIFFUSION EQUATION. See **Diffusion Equation.**

EINSTEINIUM. Chemical element symbol Es, at. no. 99, at. wt. 254 (mass number of the most stable isotope), radioactive metal of the *Actinide* series, also one of the *Transuranium* elements. Both einsteinium and fermium were formed in a thermonuclear explosion which occurred in the South Pacific in 1952. The elements were identified by scientists from the University of California's Radiation Laboratory, the Argonne National Laboratory, and the Los Alamos Scientific Laboratory. It was observed that very heavy uranium isotopes which resulted from the action of the instantaneous neutron flux on uranium (contained in the explosive device) decayed to form Es and Fm. The probable electronic configuration of Es is

$$1s^22s^22p^63s^23p^63d^{10}4s^24p^64d^{10}4f^{14}5s^25p^65d^{10}5f^{11}6s^26p^67s^2.$$

Ionic radius Es^{3+} 0.97 Å. See also **Chemical Elements.**

All known isotopes of einsteinium are radioactive. The first evidence of their existence was obtained by ion-exchange methods applied to coral rocks obtained from Eniwetok Atoll after the thermonuclear explosion. The first pure isotope found was ^{253}Es, produced by prolonged treatment of plutonium-239 with neutrons in the Arco, Idaho, Testing Reactor. The most stable is ^{254}Es, half-life 270 days, and therefore the mass number 254 is carried in the atomic weight table. Others include ^{245}Es–^{246}Es, ^{248}Es–^{252}Es, and ^{253}Es, ^{255}Es.

The ion Es^{3+} is stable. The isotopes of mass numbers 245, 252, 253 and 254 decay by alpha-particle emission; that of mass number 250 by electron capture, those of mass numbers 246, 248, 249, and 251 by both of these processes, while those of mass numbers 255 and 256 emit electrons to form the corresponding fermium isotopes.

Sufficient einsteinium, produced through intense neutron bombardments of plutonium-239 in a reactor, was not available until 1961 to allow separation of a macroscopic amount. Cunningham, Wallmann, L. Phillips, and Gatti worked with submicrogram quantities to separate a small fraction of pure einsteinium-235 compound and measure its magnetic susceptibility. Only a few hundredths of a microgram were available at that time.

Additional Reading

Armbruster, P., and G. Münzenberg: "Creating Super-heavy Elements," *Sci. Amer.*, 66–72 (May 1989).

Cunningham, B. B., Peterson, J. R., Baybarz, R. D., and T. C. Parsons: "The Absorption Spectrum of Es^{3+} in Hydrochloric Acid Solutions," *Inorg. Nucl. Chem. Lett.*, **5**, 519–523 (1967).

Cunningham, B. B., and T. C. Parsons: "Preparation and Determination of the Crystal Structure of Californium and Einsteinium Metals," *Lawrence Berkeley Laboratory Nuclear Chemistry Annual Report*, UCRL-20426, University of California, Berkeley, California, 1970.

Fields, P. R., et al.: "Additional Properties of Isotopes of Elements 99 and 100," *Phys. Rev.*, **94**, 1, 209–210 (1954).

Fisk, Z., et al.: "Heavy-Electron Metals: New Highly Correlated States of Matter," *Science*, *239*, 33–41 (1988).

Fujita, D. K., Cunningham, B. B., and T. C. Parsons: "Crystal Structures and Lattice Parameters of Einsteinium Trichloride and Einsteinium Oxychloride," *Inorg. Nucl. Chem. Lett.*, **5**, 4, 307–313 (1969).

Ghiorso, A., et al: "New Elements Einsteinium and Fermium, Atomic Numbers 99 and 100," *Phys. Rev.*, **99**, 3, 1048–1049 (1955).

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

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

Seaborg, G. T. (editor): "Transuranium Elements," Dowden, Hutchinson & Ross, Stroudsburg, Pennsylvania, 1978.

Studier, M. M., et al.: "Elements 99 and 100 from Pile-Irradiated Plutonium," *Phys. Rev.*, **93**, 6, 1428 (1954).

Thompson, S. G., Harvey, B. G., Choppin, G. R., and H. T. Seaborg: "Chemical Properties of Elements 99 and 100," *American Chemical Society Journal*, **76**, 6229–6236 (1954).

EINSTEIN LAW (Photochemistry). See **Stark-Einstein Equation.**

EINSTEIN SHIFT. According to the relativity theory, when radiation quanta leave a massive source like the sun or a star, they are retarded by the gravitational attraction and hence lose energy. This means that they lose frequency and that the wavelength λ increases. For a star of radius R and mass M, the fractional increase in wavelength is

$$\frac{\Delta}{\lambda} = \frac{G}{c^2} \cdot \frac{M}{R}$$

in which G is the gravitation constant and c is the electromagnetic constant (speed of light). The coefficient $G/c^2 = 7.414 \times 10^{-29}$ centimeters per gram. For the sun, $M = 2.3 \times 10^{33}$ grams and $R = 1.394 \times 10^{11}$ centimeters. Then $\Delta\lambda/\lambda = 1.23 \times 10^{-6}$, so that each solar spectrum line should be shifted toward the red by a little over a millionth of its own wavelength.

Measurements of this precision are hardly possible at present. But there are other stars (in particular, the white dwarfs) so massive and so dense that the shift has actually been observed and found to be of the correct order of magnitude. See also **Star.**

EJA (*Reptilia, Sauria*). The desert saw viper of Egypt, a vicious poisonous snake.

EKMAN SPIRAL (Oceanography). As originally applied by Ekman to ocean currents, a graphic representation of the way in which the theoretical wind-driven currents in the surface layers of the sea vary with depth, in an ocean assumed to be homogeneous, infinitely deep, unbounded, and having a constant eddy viscosity, over which a uniform, steady wind blows. Ekman computed that the current induced in the surface layers by the wind have the following characteristics: (1) At the very surface, the water will move at an angle of 45° *cum sole* from the wind direction; (2) in successively deeper layers, the movement will be deflected farther and farther *cum sole* from the wind direction, and the speed will decrease; and (3) a hodograph of the velocity vectors would form a spiral descending into the water and decreasing in amplitude exponentially with depth.

ELASTIC CONSTANTS AND MODULI. These constants occur in relationships between the components of strain and stress in a medium which obeys Hooke's law. Under such conditions the strain components may be expressed as functions of the stress components by linear equations of the type:

$$e_{xx} = s_{11}X_x + s_{12}Y_y + s_{13}Z_z + s_{14}Y_z + s_{15}Z_x + s_{16}X_y$$

where e_{xx} is the strain component along x-axis, the capitalized quantities are the stress components (the stress component X_x, for example, representing a force applied in the x-direction to a unit area of a plane whose normal lies in the x-direction) and the s terms are called the elastic constants or compliance constants. Also, the stress components may be expressed as functions of the strain components by linear equations of the type:

$$X_x = c_{11}e_{xx} + c_{12}e_{yy} + c_{13}e_{zz} + c_{14}e_{yz} + c_{15}e_{zx} + c_{16}e_{xy}$$

where the c terms are called the stiffness constants or elastic moduli.

For the special moduli used in engineering, see **Elasticity.**

ELASTIC CURVE. The curve of the neutral surface of a structural member subjected to loads which cause bending is called the elastic curve. The ordinates between this curve and the original position of the neutral surface represent the deflections due to bending.

ELASTIC DEFORMATION. See **Resilience.**

ELASTICITY. The property whereby a body, when deformed, automatically recovers its normal configuration as the deforming forces are removed. Each of its several types is probably due to the action of intermolecular forces which are in equilibrium only for certain configurations.

Deformation or, more briefly, strain is of various kinds; in each case its measure is a certain abstract ratio. For example, the elongation of a rod under tension is expressed as the ratio of the increase in length to the unstretched length. Linear compression is the reverse of elongation. They are both accompanied by a fractional change in diameter, the ratio of which to the elongation is called the Poisson ratio. Shear is a strain involving change of shape, such that an imaginary cube traced in the unstrained material becomes a rhombic prism. The measure of shear is the tangent of the angle through which the oblique edges have been made to depart from their original perpendicular direction. Volume strain is the ratio of a decrease in volume to the normal volume. Flexure, or bending, and torsion, or twisting, are combinations of these more elementary strains. A straight rod bent into a plane curve undergoes elongation on the convex side and linear compression on the concave side, while there is an intermediate neutral layer which suffers neither.

In tests such as the tensile test it is necessary to differentiate between two basic ways of measuring strain. "Conventional strain" is the in-

$$\ln \frac{l}{l_0}$$

crease in length of the gage section divided by the original gage length, whereas the "true strain" is the natural logarithm
where l_0 is the original value and l the instantaneous value of the length of the gage section.

For every strain, there arises, in an elastic substance, a corresponding stress, which represents the tendency of the substance to recover its normal condition. Stress is expressed in units of force per unit area. Tensile stress, for example, is the ratio of the force of tension to the normal cross section of the rod subjected to it. Shearing stress is the force tending to push one layer of the material past the adjacent layer, per unit area of the layers. Pressure, expressed in like units, is the stress corresponding to volume compression, etc.

For each type of strain and stress there is a modulus, which is the ratio of the stress to the corresponding strain. In the case of elongation or linear compression, it is commonly called Young's modulus; we also have the bulk modulus and the shear modulus of rigidity. See accompanying table.

NOMINAL MODULUS VALUES OF REPRESENTATIVE METALS
(values in pounds per square inch)

Material	Young's Modulus	Rigidity of Shear Modulus
Magnesium	6.5×10^6	2.4×10^6
Aluminum	10.2×10^6	3.6×10^6
Copper	14.5×10^6	6.1×10^6
Steel	30×10^6	11.6×10^6

In engineering design, Young's modulus is used for tension and compression and the rigidity modulus for shear, as in torsion springs. According to Hooke's law, the stress set up within an elastic body is proportional to the strain to which the body is subjected by the applied load.

Theory of Elasticity. The fundamental quantities in elasticity are second-order tensors, or dyadics: the deformation is represented by the *strain dyadic*, and the internal forces are represented by the *stress dyadic*. The physical constitution of the deformable body determines the relation between the strain dyadic and the stress dyadic, which relation is, in the infinitesimal theory, assumed to be linear and homogeneous. While for anisotropic bodies this relation may involve as much as 21 independent constants, in the case of isotropic bodies, the number of elastic constants is reduced to two.

Let $\mathbf{s(r)}$ be the displacement vector, due to the deformation, of a particle that before the deformation was situated at point P having $\mathbf{r}$ as position vector with respect to some arbitrary origin. A neighboring point Q, whose position vector was $\mathbf{r} + d\mathbf{r}$ before the deformation, will suffer a displacement $\mathbf{s}(\mathbf{r} + d\mathbf{r})$ which will differ from $\mathbf{s(r)}$ by the quantity

$$d\mathbf{s} = d\mathbf{r} \cdot \nabla \mathbf{s}$$

The hypothesis of small deformations means that ds, the change in the displacement vector when we go from P to the neighboring point Q, is very small compared to $d\mathbf{r}$, the position vector of Q relative to P. Consequently, the scalar components of the dyadic $\nabla\mathbf{s}$ are all very small compared to unity. The geometrical meaning of the dyadic $\nabla\mathbf{s}$ is obtained by separating it into its symmetric part $\mathbf{S} = \frac{1}{2}(\nabla\mathbf{s} + \mathbf{s}\nabla)$ and its antisymmetric part $\mathbf{R} = -\frac{1}{2}\mathbf{1} \times (\nabla \times \mathbf{s})$, where $\mathbf{1}$ is the unity dyadic.

The antisymmetric part is interpreted as follows: if at some point M the symmetric part vanishes, then we have for the neighborhood of M the relation

$$d\mathbf{s} = d\mathbf{r} \cdot \mathbf{R}_M = \omega_M \times d\mathbf{r}$$

where $\omega_M = \frac{1}{2}(\nabla \times \mathbf{s})_M$ is an infinitesimal vector. This means that the neighborhood of point M undergoes an infinitesimal rigid rotation, without any change in shape or size. Consequently, the deformation is represented by the symmetric part $\mathbf{S}$, which is called the *strain dyadic*.

In a Cartesian orthonormal basis, in which we have $\mathbf{r} = \Sigma_{i=1}^{3} x_i\mathbf{a}_i$, we write $\mathbf{s} = \Sigma_{i=1}^{3} s_i\mathbf{a}_i$, and obtain

$$\mathbf{S} = \sum_{i,j=1}^{3} \mathbf{a}_i\mathbf{a}_j S_{ij}$$

where

$$S_{ij} = \frac{1}{2}\left[\frac{\partial}{\partial x_i} s_j + \frac{\partial}{\partial x_j} s_i\right].$$

The diagonal components S_{11}, S_{22}, and S_{33} are the coefficients of linear extension in the directions $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$, respectively, while the non-diagonal components $S_{12} = S_{21}$, $S_{13} = S_{31}$, and $S_{23} = S_{32}$ are called shear strains. For instance, $2S_{12}$ is the change in the angle of the dihedron formed by the planes that before the deformation were respectively normal to the directions $\mathbf{a}_1$ and $\mathbf{a}_2$. The shear strains are not essential for the complete representation of a deformation since they can be made to vanish by expressing $\mathbf{S}$ in the basis of its principal axes.

If an infinitesimal element of the body occupies the volume dV before the deformation and the volume dV' after, the relative increase of volume, or volumetric dilatation, is given by

$$\frac{dV' - dV}{dV} = S_{11} + S_{22} + S_{33} = |\mathbf{S}| = \nabla \cdot \mathbf{s}$$

The forces applied to a finite deformable body are either body forces acting on every volume element dV and represented by the notation $dV\,\mathbf{F} = dV\,\rho\mathbf{K}$, where $\mathbf{F}$ is the force per unit volume, $\mathbf{K}$ is the force per unit mass, and ρ is the density, or surface forces acting on every element dS of the bounding surface and represented by $dS\,\mathbf{T}$, where $\mathbf{T}$ is the surface stress, or surface force per unit area. The effect of these applied forces is transmitted throughout the body, so that through any surface element inside the body, there is a force exerted by the matter on one side of the element upon the matter on the other side. Such forces are called internal stresses and are defined as follows: let dS be a surface element completely inside the body, and let us choose arbitrarily the positive sense of the normal $\mathbf{n}$ to this surface element; this defines for dS a positive side, the one containing $\mathbf{n}$, and a negative side. Then $\mathbf{T}_n$, the stress vector on the positive side of dS is defined as a vector such that $dS\,\mathbf{T}_\mathbf{n}$ is the surface force on the positive side of dS—i.e., the resultant of all the forces exerted through dS by the matter on the positive side of dS upon the matter on the negative side. In general there is a normal component $\mathbf{T}_\mathbf{n}\cdot\mathbf{nn}$, which is a pressure or a traction depending upon whether the sign of $\mathbf{T}_\mathbf{n}\cdot\mathbf{n}$ is negative or positive, and a tangent component $\mathbf{n} \times \mathbf{T}_\mathbf{n} \times \mathbf{n}$ called the shear stress. The value of the stress vector $\mathbf{T}_\mathbf{n}$ depends upon the orientation of the normal $\mathbf{n}$, so that we can characterize the state of stress at a point by defining the *stress dyadic* $\mathbf{T}$ through the relation

$$\mathbf{T}_\mathbf{n} = \mathbf{n} \cdot \mathbf{T}$$

The mechanical equilibrium conditions applied to an arbitrary volume V, bounded by the closed surface S, and completely inside the deformable body give

$$\int_V dV\,\mathbf{F} + \int_S dS\,\mathbf{n} \cdot \mathbf{T} = 0$$

and

$$\int_V dV\,\mathbf{r} \times \mathbf{F} + \int_S dS\,\mathbf{r} \times (\mathbf{n} \cdot \mathbf{T}) = 0$$

By the use of the divergence theorem, the first condition gives the equation

$$\nabla \cdot \mathbf{T} + \mathbf{F} = 0$$

at any point inside the body, and the second condition implies that $\mathbf{T}$ is a symmetric dyadic. On the external surface of the body, we have usually to fulfill the boundary condition

$$\mathbf{n} \cdot \mathbf{T} = \mathbf{T}$$

where $\mathbf{T}$ is the applied external force per unit area. Other boundary conditions can also be met, such that the value of the displacement be prescribed.

For infinitesimal deformations, we assume that the relation between strain and stress is expressed by Hooke's law: the deformation is proportional to the applied force. For isotropic bodies, this linear relation is

$$\mathbf{S} = \frac{1}{E}\,[(1 + \nu)\mathbf{T} - \nu|\mathbf{T}|\mathbf{1}]$$

where E is Young's modulus and ν is Poisson's ratio. These two elastic constants can be defined by considering the stretching of a cylindrical bar by normal traction forces uniformly distributed on the end sections; then we have

$$\text{Young's modulus} = \frac{\text{Normal traction force per unit cross sectional area}}{\text{Relative longitudinal extension}}$$

and

$$\text{Poisson's ratio} = \frac{\text{Relative lateral contraction}}{\text{Relative longitudinal extension}}$$

We can also write

$$\mathbf{T} = 2\mu\mathbf{S} + \lambda|\mathbf{S}|\mathbf{1}$$

where $\mu = E/2(1 + \nu)$ and $\lambda = \nu E/(1 + \nu)(1 - 2\nu)$ are Lamé's constants. μ is the rigidity modulus, the only constant necessary when the volumetric dilatation vanishes everywhere.

Substituting the preceding relation into the equilibrium equations, we transform them into

$$2\mu\nabla \cdot \mathbf{S} + \lambda\nabla|\mathbf{S}| + \mathbf{F} = 0 \text{ inside the body}$$

and

$$2\mu\mathbf{n} \cdot \mathbf{S} + \lambda\mathbf{n}|\mathbf{S}| = \mathbf{T} \text{ on the bounding surface.}$$

These vector relations are not sufficient for the complete determination of the symmetric dyadic $\mathbf{S}$. To insure that a solution of the above equations corresponds to a possible displacement vector $\mathbf{s}$, we must be able to integrate the relation

$$\mathbf{S} = \tfrac{1}{2}(\nabla\mathbf{s} + \mathbf{s}\nabla)$$

i.e., from a given expression for $\mathbf{S}$ obtain the value of $\mathbf{s}$. From the vanishing of the curl of a gradient, it is easily seen that this integrability condition, also called the compatibility equation, is

$$\nabla \times \mathbf{S} \times \nabla = 0$$

By elimination of the vector products, we obtain the equivalent form

$$\nabla\nabla \cdot \mathbf{S} + \nabla \cdot \mathbf{S}\nabla - \nabla\nabla|\mathbf{S}| - \nabla \cdot \nabla\mathbf{S} = 0$$

Using the stress-strain relation and the equilibrium conditions, we obtain the Beltrami-Michell form of the compatibility equation:

$$\nabla \cdot \nabla \mathbf{T} + \frac{1}{1+\nu} \nabla\nabla |\mathbf{T}| = -\frac{\nu}{1-\nu} \nabla \cdot \mathbf{F1} - (\nabla\mathbf{F} + \mathbf{F}\nabla)$$

Finally, by expressing the strain dyadic in terms of the displacement vector, we obtain Navier's form of the equilibrium equations:

$$\mu\nabla \cdot \nabla\mathbf{s} + (\lambda + \mu)\nabla\nabla \cdot \mathbf{s} + \mathbf{F} = 0$$

inside the body and

$$\lambda n\nabla \cdot \mathbf{s} + 2\mu\mathbf{n} \cdot \nabla\mathbf{s} + \mu\mathbf{n} \times (\nabla \times \mathbf{s}) = \mathbf{T}$$

on the bounding surface. Dealing here directly with the displacement vector, there is no need of considering the compatibility equation.

The propagation equation for elastic disturbances is obtained by adding the inertia force to the body force. We get then

$$\mu\nabla \cdot \nabla\mathbf{s} + (\lambda + \mu)\nabla\nabla \cdot \mathbf{s} + \rho\mathbf{K} = \rho\frac{\partial^2}{\partial t^2}\mathbf{s}$$

inside the body. The stress-strain relation and the boundary conditions are not affected, but we generally have to take into account initial conditions.

The energy density u, or energy per unit volume, is given by

$$u = \frac{1}{2}\,\mathbf{SS:TT} + \frac{1}{2}\rho\frac{\partial\mathbf{s}}{\partial t} \cdot \frac{\partial\mathbf{s}}{\partial t}$$

where the first term is potential, or strain energy, and the second term is kinetic energy. The energy flux density vector

$$\mathbf{S} = -\frac{\partial\mathbf{s}}{\partial t} \cdot$$

is a vector such that $dS\,\mathbf{n} \cdot \mathbf{S}$ gives the quantity of energy that flows per unit time through the surface element dS in the positive direction of $\mathbf{n}$, the normal to dS. At any point the energy continuity equation

$$\frac{\partial u}{\partial t} + \nabla \cdot \mathbf{S} - \rho\frac{\partial\mathbf{s}}{\partial t} \cdot \mathbf{K} = 0$$

expresses the conservation of mechanical energy.

ELASTICITY (Modulus). See **Modulus of Elasticity.**

ELASTIC LIMIT. The maximum unit stress which can be obtained in a structural material without causing a permanent deformation is called the elastic limit.

ELASTIC REBOUND THEORY. See **Earth Tectonics and Earthquakes.**

ELASTOMERS. Of natural or synthetic origin, an elastomer is a polymer possessing elastic (rubbery) properties. A polymer is a substance consisting of molecules which are, in the most part, multiples of low-molecular-weight units, or monomers. As an example, isoprene (2-methylbutadiene-1,3) is C_5H_8, whereas polyisoprene is $(C_5H_8)_x$, where $x \geq 2$ and normally is from 1,000 to 10,000 for rubbers. Although they differ in composition from natural rubber, many of these high-molecular-weight materials are termed *synthetic rubbers*. See also **Rubber (Natural).**

The serious development of synthetic rubbers commenced in the late 1930s and early 1940s, accelerated by a cutoff of supplies of natural rubber because of political turmoil and war. Synthetic rubbers fall into two major classifications: (1) general-purpose rubbers, the major volume of which is nevertheless used for tire production; and (2) specialty rubbers that essentially find little use in tires, but that are important for a number of other categories. Synthetic rubbers have not replaced natural rubber for numerous uses. For large, heavy-duty truck and bus tires, natural rubber tends to run considerably cooler and wears better than a blend of natural and synthetic rubbers. On the other hand, a tire tread made of a blend of styrene-butadiene (SBR) and butadiene rubber (polybutadiene) wears longer than natural rubber in conventional automobile, usage, where lower temperatures can be maintained.

Styrene-Butadiene (SBR) Rubbers. This series of rubbers includes monomer ratios up to about 50% styrene. The addition of more than 50% styrene makes the materials more like plastic than rubber. The most commonly used SBR rubbers contain about 25% styrene, which is polymerized in emulsion systems at 5–10°C. Most SBR goes into tires, but the type for the tread differs from that of the sidewall or carcass. SBRs for adhesives, shoe soles, and other products also differ. The formulation permits vast varieties of end-products. Among the processing variables that can be manipulated to provide different end-characteristics are temperature, viscosity, use of different emulsifiers and solvents, use of different antioxidants for stabilization, different oils, carbon blacks, and coagulation techniques.

Initial processes for emulsion (E-SBR) called for polymerization at 50°C. However, it was found later that cold processing produced a rubber particularly good for tires. This is sometimes referred to as *cold SBR*. The overall process for making SBR is shown in the accompanying figure. Typical formulations are:

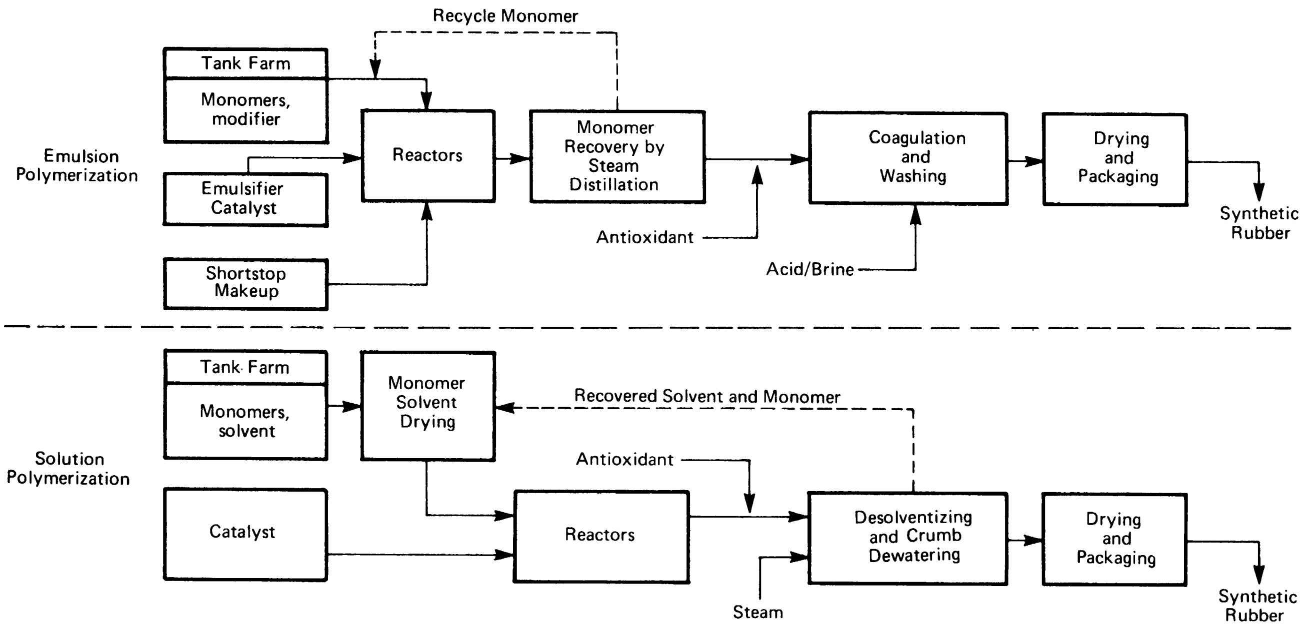
	Parts by Weight	
	Hot SBR	*Cold SBR*
Deionized water	180	200
Fatty acid soap	5	4.5
Styrene	25	28
t-Dodecyl-mercaptan	0.35	0.2
Butadiene	75	72
Potassium persulfate	0.3	0
Redox initiator	0	small quantity
Temperature	50°C	5°C
Time	12 hours	8 hours
Conversion	75%	60%

Polymerization of emulsion SBR is started by free radicals generated by the redox system in cold SBR and by persulfate or other initiator in hot SBR. The initiators are not involved in the molecular structure of the polymers. Almost all molecules are terminated by fragments of the chain transfer agent (a mercaptan). Schematically, the molecules are RSM_nH, where RS is the $C_{12}H_{25}S$ part of a dodecyl mercaptan molecule; M is the monomer involved; n is the degree of polymerization, and H is a hydrogen atom formerly attached to the sulfur of a mercaptan. In the case of free-radical-initiated polymerization of butadiene, by itself to form homopolymers or with other monomers for form copolymers, the butadiene will be about 18%; 16% *cis*-1,4; and 66% *trans*-1,4.

Considerable quantities of SBR latex are used in the manufacture of foam rubber, adhesives, fabric treating, and paints. The solids content of latices runs from 50% to 65–70%.

Solution (S-SBR) consists of styrene butadiene copolymers prepared in solution. A wide range of styrene-butadiene ratios and molecular structures is possible. Copolymers with no chemically detectable blocks of polystyrene constitute a distinct class of solution SBRs and are most like styrene-butadiene copolymers made by emulsion processes. Solution SBRs with terminal blocks of polystyrene (S-B-S) have the properties of self-cured elastomers. They are processed like thermoplastics and do not require vulcanization. Lithium alkyls are used as the catalyst.

Stereospecific solution polymerization has been emphasized since the discovery of the complex coordination catalysts that yield polymers of butadiene and isoprene having highly ordered microstructures. The catalysts used are usually mixtures of organometallic and transition-metal compounds. An example of one of these polymers is *cis*-1,4-polybutadiene, the bulk of which is used in tires. However, it must be blended with other materials because of its poor processability and traction.



Synthetic rubber production process.

Butyl Rubber. Known as IIR, butyl rubber is a copolymer of isobutylene and isoprene. The elastomers contain only 0.5–2.5 mole % of isoprene. This is introduced to effect sufficient unsaturation to make the rubber vulcanizable. Polymerizations are usually carried out at low temperature (−80 to −100°C) with methyl chloride as solvent. Anhydrous aluminum chloride and a trace of water serve as catalyst.

Butyl is one of the earlier synthetic rubbers. However, it lost favor after the development of SBR. Butyl rubber is incompatible with natural rubber and is difficult to cure. Chlorobutyl rubbers, a much more recent development and containing up to 1.3% chlorine, apparently do not exhibit these former problems. Major uses for butyl rubber have been inner tubes for tires. The appearance of tubeless tires, however, depressed this market. Butyl rubber makes excellent motor mounts because of its high energy absorption and low rebound. Essentially free of double-bonds, butyl rubber has a high resistance to aging, attractive properties for use in curing bags for tire production and in outside coating materials.

Acrylonitrile-Butadiene Rubbers (NBR). Except for the monomers used, the production of NBRs is quite similar to that described for the SBRs. The NBR family is sometimes referred to as the *nitrile rubbers.* The acrylonitrile-butadiene ratios cover a wide range from 15:85 to 50:50. NBRs are noted for their solvent resistance, increasing with the acrylonitrile content. Thus, they are used for gaskets and oil and gasoline hoses, solvent-resistant electrical insulation, and food-wrapping films. Nitrile latices also are used in treating fabrics for dry-cleaning durability. Because the NBRs become quite inflexible (stiff) at low temperatures (actually brittle at about −20°C), they are blended with polyvinyl chloride for some applications.

Neoprene. This family of dry rubbers and latices was introduced in 1932, called *Duprene* by Du Pont at that time. The material is made by the free-radical-initiated polymerization of chloroprene in emulsion systems. As with most synthetic rubbers, a variety of neoprenes is made possible by variation of the polymerization conditions and ingredients. Neoprene is particularly good for its fire-retardant, solvent-resistant, and high-temperature stability properties. The chlorine in each segment deactivates the adjoining carbon-carbon double-bond, thus making it less sensitive to oxidative attack. Metal oxides, such as zinc oxide and magnesium, serve as curatives rather than sulfur.

Polyurethanes. Polyethylene in solution is treated with chlorine and sulfur dioxide to introduce approximately 1.3% sulfur and 29% chlorine into the polymer. Most of the chlorine is attached directly to the carbon atoms in the backbone of the polymer. The remainder is in the form of sulfuryl chloride groups, $\cdot SO_2Cl$, through which crosslinking occurs in the curing step with metal oxides. The material has good oxidation and ozone resistance and thus overall excellent weather resistance. Calendered stocks are used for lining ditches and ponds, for example.

Thiokol® Rubbers. These are polysulfide rubbers and are prepared by the condensation polymerization of sodium polysulfides with a dichloro (sometimes blended with a trichloro) organic compound. Type A, the first family of rubbers, was made from Na_2S_4 and ethylene dichloride. Thiokols are known for high resistance to organic solvents.

Polyacrylate Elastomers. These are made in emulsion or suspension systems involving the copolymerization of ethyl acrylate with the acrylate esters of higher-molecular-weight alcohols. These materials have excellent solvent-resistant properties and stability at elevated temperatures. A major use is for automatic-transmission gaskets for automobiles.

Silicone Elastomers. These materials have alternating Si and O atoms for a backbone and the members differ mainly in the nature of the organic substituents on the Si atoms and the degree of polymerization. Because of the absence of double-bonds in the backbone, the numerous forms of stereoisomers found in unsaturated hydrocarbon rubbers do not have counterparts in the silicone rubbers. The chemical combination of organic and inorganic materials gives the silicone rubbers useful properties over a wide temperature range (−70 to + 225°C). These rubbers are well known for excellent dielectric stability and high resistance to weathering, oils and chemicals.

Fluoroelastomers. These materials are prepared by emulsion copolymerization of perfluoropropylene and vinylidene fluoride; or of chlorotrifluoroethylene and vinylidene fluoride. Also there are fluorosilicones in this family. Useful at temperatures up to and over 300°C, the fluoroelastomers have excellent resistance to aromatic solvents, acids, and alkalies. They are also among the more costly of the available commercial elastomers.

Ethylene-Propylene Elastomers. Known as EPR, this material is of limited use because it cannot be vulcanized in readily available systems. However, the rubbers are made from low-cost monomers, have good mechanical and elastic properties, and outstanding resistance to ozone, heat, and chemical attack. They remain flexible to very low temperatures (brittle point about −95°C). They are superior to butyl rubber in dynamic resilience.

RECORD ELDER TREES AND VIBURNUMS IN THE UNITED STATES[1]

Specimen	Circumference[2] (Inches)	Circumference[2] (Centimeters)	Height (Feet)	Height (Meters)	Spread (Feet)	Spread (Meters)	Location
BLACKHAWS							
Viburnum prunifolium (1966)	39	99	29	8.8	30	9.1	Pennsylvania
Rusty (1961) *Viburnum rufidulum*	47	119	25	7.6	30	9.1	Arkansas
ELDERS							
American (1987) *Sambucus canadensis var. canadensis*	38	97	16	4.9	22	6.7	Virginia
Blackbead (1972) *Sambucus melanocarpa*	39	99	42	12.8	30	9.1	Oregon
Blue (1979) *Sambucus cerulea*	137	348	40	12.2	36	11.0	California
Florida (1972) *Sambucus canadensis var. laciniata*	34	86	20	6.1	14	4.3	Florida
Mexican (1981) *Sambucus mexicana*	108	274	31	9.4	32	9.8	New Mexico
Pacific Red (1989) *Sambucus callicarpa*	90	229	30	9.1	44	13.4	Oregon
Scarlet (1989) *Sambucus pubens*	18	48	27	4.9	13	4.0	Michigan
VIBURNUMS							
American Cranberrybush (1985) *Viburnum triolobum*	10	25	32	9.8	31	9.4	Michigan
Possumhaw (1972) *Viburnum nudum*	12	30	26	7.9	9	2.7	Florida
Walter (1976) *Viburnum obovatum*	22	56	30	9.1	23	7.0	Florida

[1]From the "National Register of Big Trees," The American Forestry Association (by permission).
[2]At 4.5 feet (1.4 meters).

ELDER TREES AND VIBURNUMS. Of the family *Caprifoliaceae* (elder or honeysuckle family), the elder is a small tree, frequently a shrub, which is found in Europe and North America. The box elder (or ash-leaved maple) is a member of the family *Aceraceae* (maple family) and is not directly related to the other elders. Elder trees are of the genus *Sambucus*. These plants are deciduous, considered hardy, with odd, pinnately compounded leaves. There are from 5 to 11 ovate, pointed, finely-toothed leaflets. The flower is small and may be pink or creamy white in color, occurring in flat clusters. The fruit is a small berrylike drupe and may range in color from red, through blues and purples, and black, depending upon the species. There are about 12 species commonly found in North America. The more important include: American elder (*Sambucus canadensis*), which has black-to-purple-to-blue berries, well known as the source of elderberry wine; the blueberry elder (*S. glauca*), which has a blue fruit; the Mexican (*S. Mexicana*) and the velvet elder (*S. velutina*), both of which have black fruits; and the Pacific red elder (*S. callicarpa*), which has a red fruit. There are also some essentially local species, including the blackhead elder (*S. melanocarpa*), and the Florida elder (*S. simpsonii*). See accompanying table for record elder trees.

The American elder is found from Nova Scotia and New Brunswick westward across Canada to Manitoba and south and westward in the United States through Kansas and into Texas and Arizona. The tree tends to become larger as it is found in the west. The elder can be found in the Allegheny Mountains to an elevation of about 3500 feet (1070 meters). It thrives in rich moist soil.

The common elder of Europe, north Africa, and western Asia is the *S. nigra*, which has yellowish flowers and lustrous black fruit. The pithy new wood of this shrub or tree is known for its rather disagreeable odor, which is alleged to even repel flies and other insects.

Closely associated with the elders and also members of the elder family are the viburnums, which also are deciduous, but sometimes evergreen shrubs and trees. These plants have fragrant, clustered white flowers and fruits of fleshy consistency and bright colors. The *Viburnium prunifolium* is commonly referred to as the blackhaw or stag-bush. Usually an erect bush, but frequently a tree ranging in height from 10 to close to 30 feet (3 to 9 meters), the plant normally will have a trunk diameter up to 10 inches. The bark is rough and gray-brown and in earlier years was used as a tonic and medicinal. The leaves are dark green, from 1 to 3 inches (2.5 to 7.6 centimeters) in length. The flowers are quite small and white and occur in large clusters approaching 5 inches (12.7 centimeters) in breadth. The fruit is bluish, edible, and sweet. The southern or rusty blackhaw (*V. rufidulum*) is a somewhat smaller species, but quite similar to the stag-bush. Record blackhaws also are listed on the accompanying table.

There are numerous species of the viburnums, just a few of which include: The wayfaring tree (*V. lantana*), very popular in Europe as a hedgerow plant; *V. opulus*, also known in Europe and north Africa as the Guelder rose or cranberry shrub, and particularly well adapted to wet areas. (Incidentally, this shrub is not related to the common cranberry, which is of the family *Ericaceae* (heather family).); the *V. alnifolium*, also known as hobble bush, witch hobble, and wayfaring tree; *V. opulus* var. *americanum*, commonly known as the pimbina, crampbark, cranberry tree, or highbush cranberry (these berries are sometimes used as substitutes for common cranberries); *V. pauciflorum*, known as the squashberry bush or squashberry pimbina; *V. acerifolium*, the dockmackie or arrowwood; *V. pubescens*, the downy arrowwood; *V. molle*, the soft-leaved arrowwood; *V. venosum*, similar to *V. molle*; *V. dentatum*, the arrowwood shrub; *V. cassinoides*, the withe-rod Appalachian tea shrub; *V. nudum*, the naked withe-rod; and *V. lentago*, the nannyberry, sheepberry, or wild raisin tree.

ELECTRET. A permanently-polarized piece of dielectric material; the analog of a magnet. Barium titanate ceramics, preferably containing a small percentage of lead titanate, can be polarized by cooling from a temperature above the Curie point in an applied electric field. Electrets are also produced by solidification of mixtures of certain organic waxes in a strong electric field.

ELECTRICAL CONDUCTIVITY. The measure of a material's ability to carry an electric current. An electric conductor is a material which, when placed between terminals having a difference of electrical potential, will readily permit the passage of an electric current. Different materials have different degrees of conductivity, and their effectiveness in this respect is computed as the conductivity. The best conductors are the metals, such as silver, copper, aluminum, platinum, and mercury, but nonmetallic substances such as carbon, saline solutions, and moist earth also are sufficiently conductive so that this property becomes of significance under certain circumstances. By virtue of their cost-conductivity characteristic, copper and aluminum are widely used conductors. They will usually be found as wires or buses. Copper is used more commonly than aluminum, the use of which is preferred for high-voltage transmission lines, where its lighter weight is of definite advantage. Steel as a conductor is inferior to the other two materials mentioned. but its greater strength and resistance to wear have led to its adoption as a conductor of special purposes, such as that of power rail service on electrified railways, and as an inner core of copper or aluminum cables. The resistance of a conductor is its resistivity multiplied by its length and divided by its cross-sectional area.

Commonly Used Electrical Conductors. For practical comparative purposes, the commonly used metals are compared with the International Annealed Copper Standard (IACS). A value of 100% conductivity is assigned to annealed copper. The standard may be expressed in terms of mass resistivity as 0.15328 ohm-grams/square meter, or the resistance of a uniform round wire 1 meter long weighing 1 gram at 20°C. Useful equivalent expressions of the annealed copper standard in terms of various units of mass resistivity and volume resistivity, include:

Value	Units at 20°C
0.15328	ohm-grams/square meter
875.20	ohm-pounds/square mile
1.7241	microohm-centimeters
0.67879	microohm-inches
10.371	ohm-circular mils/foot
0.017241	ohm-square millimeters/meter

All of the foregoing values are equivalent to $\frac{1}{58}$ ohm-square millimeters/meter. Thus volume conductivity can be expressed as 58 mho-square millimeters/meter at 20°C. See accompanying table.

CONDUCTORS—ELECTRICAL CONDUCTIVITY AND RESISTIVITY[a]

Material	% IACS (Volume) 20°C	Volume Resistivity (Microohm-Centimeters) 20°C
Annealed copper	100.00	1.7241
Copper alloy (B187-62)	98.40	1.7521
Copper alloy (B188-61)	97.80	1.7629
Copper alloy (B47-64; B116-64)	97.16	1.7745
Copper alloy (B355-65T, nickel coated)	96.00	1.7960
Copper alloy (B246-64, tinned hard)	92.72	1.8595
Copper alloy (B355-65T, nickel coated, soft)	88.0	1.9592
Aluminum	64.94	2.6548
Aluminum alloy (B233-64)	61.5	2.8035
Aluminum alloy (B317-64)	59.0	2.9222
Aluminum alloy (B396-63T)	52.5	3.2839
Copper-clad steel	40	4.3971
Aluminum-clad steel	20	8.4805
Iron (pure)		9.78
Commercial galvanized iron		16–20
Silver	108.4	1.59
Gold		2.3

[a]Representative metals and alloys. There are hundreds of metal combinations used. Frequently, electrical conductivity is not the sole criterion in selecting a conductor. Other factors include density (weight), strength, ductility, and corrosion and abrasion resistance.

Fundamentals of Electrical Conductivity. The conductivity σ of an isotropic material in a steady direct-current electric field E is defined by

$$j = \sigma E$$

where j is the current density (charge transported per unit time across unit area perpendicular to the current flow). In meters-kilograms-seconds (mks) units, j is measured in amperes per square meter and E in volts per meter, so that σ is in (ohm-meters)$^{-1}$, or mhos per meter. The reciprocal of the conductivity is the electrical resistivity, $\rho = 1/\sigma$. The electrical resistance R of a sample is the ratio of the potential drop across the sample to the total current through the sample. The resistance of a cylindrical sample (with the current flow parallel to the axis of the cylinder) is

$$R = \rho l/A$$

where l is the length of the cylinder and A is its cross-sectional area. If the dimensions are measured in meters and ρ is in mks units, the resistance is given in ohms. The reciprocal of the resistance is the electrical conductance.

Many homogeneous solids and liquids obey Ohm's law for sufficiently small electric fields. Ohm's law states that the current through a sample is proportional to the potential drop across the sample, and thus R, ρ, and σ are independent of the impressed electric field. Samples which do not obey Ohm's law are called *nonlinear*; gases fall into this category, as do many important circuit elements, such as transistors and vacuum tubes. Most metals obey Ohm's law for field up to at least 10^8 volts/ meter, though under certain conditions, semiconductors have shown deviations from Ohm's law for fields as low as 10^{-2} volts/meter.

The existence of a finite resistance means that the energy delivered bu the electric field to the current carriers is dissipated, being converted to heat (mostly energy of atomic vibrations). The rate of dissipation per unit volume is given by

$$\tfrac{1}{2}\rho j^2 = \tfrac{1}{2}\sigma E^2$$

and is called the Joule heat.

In some solids, the conductivity is *anisotropic*, i.e., the magnitude of j depends upon the direction of E as well as its magnitude, and j and E are not necessarily parallel. If Ohm's law is obeyed, it becomes

$$\mathbf{j} = \boldsymbol{\sigma E}$$

where σ is a second rank tensor. Anisotropic solids include single crystals of materials which do not have cubic crystal structures, and polycrystalline aggregates of such materials in which there exists some preferred orientation such as can be produced by extrusion or cold rolling.

If the applied electric field varies sinusoidally in time (alternating current), the conductivity generally depends upon the applied frequency v. Appreciable deviations from the dc value may appear for microwave frequencies or greater. In the microwave and optical range, it is common to identify $-\sigma/2\pi\epsilon_0\, v$ as the imaginary part of the *dielectric coefficient*, where ϵ_0 is the permittivity of free space $\frac{1}{4}\pi\epsilon_0 = 9 \times 10^9$ newton meter^2/coulomb^2. Thus, σ is closely related to the absorption of electromagnetic energy.

Solids are usually classified as metals, semiconductors, or insulators. Metals are characterized by an increasing resistivity with increasing temperature. Resistivities of metals at room temperature range from about 10^{-8} to 10^{-6} ohm-meters. Behavior at extremely low temperatures is covered under **Superconductivity.** Semiconductors are characterized by a decreasing resistivity with increasing temperature (impure semiconductors may show this behavior only at high temperatures). Resistivities of semiconductors at room temperature range from about

10^{-5} to 10^{+7} ohm-meters. Insulators share the same temperature behavior of resistivity as semiconductors, so the difference between the two classes of materials is a simple one of degree only. Generally, materials with room-temperature resistivities greater than 10^7 ohm-centimeters are called insulators. Resistivities as high as 10^{18} ohm-meters have been observed. There are some materials intermediate between metals and semiconductors. For example, the resistivity of manufactured carbon decreases with increasing temperature at low temperatures, and increases at high temperatures. The room-temperature resistivity is also intermediate, varying from 10^{-5} to 10^{-4} ohm-meters, depending upon the conditions of manufacture.

Except in the case of some insulators, the current in solids is carried by electrons. Quantum mechanics has shown that not all the electrons in a solid are free to carry current. The conductivity depends upon the number of free carriers and their ease of motion. The latter factor is expressed by the mobility μ which is defined by

$$v_d = \mu E$$

where v_d is the drift velocity (average velocity of the free carriers produced by the action of the electric field). The drift velocity is usually very much smaller than a typical carrier velocity v (the average of the magnitude of the carrier velocities). The conductivity is then given by

$$\sigma = ne\mu$$

where n is the density of free carriers (in meters^{-3}), e is the magnitude of the electronic charge (1.6×10^{-19} coulomb), and μ is in square meters per volt per second. If more than one group of carriers is present, the total conductivity is the sum of contributions from each group. Quantum mechanics has also shown that the mobility in a pure, perfectly regular crystal would be unlimited. The mobility is limited by the scattering of the carriers by deviations, such as the thermal vibrations of the atoms, impurity atoms, or irregularities in the crystal structure, such as vacancies and dislocations. A simple approximate theory of the mobility allows it to be expressed as

$$\mu = e\tau/4\pi\epsilon_0 m^* = e\lambda/4\pi\epsilon_0 mv$$

where τ is the relaxation time (roughly, the average time between the collisions which a carrier makes with the scattering centers), λ is the mean free path of the carriers, and m^* is the effective mass of the carrier (a concept which comes from the theory of energy bands in solids). If more than one scattering process is involved, the reciprocal of the relaxation time (scattering rate) is approximately the sum of the contributions from each process. At room temperature, mobilities range from very small values (such as 10^{-4} meter^2/volt-second) to 10 meter^2/ volt-second; relaxation times range from about 10^{-14} second to about 10^{-12} second, and mean free paths range from about 10^{-9} meter to about 10^{-6} meter. One of the early triumphs of quantum mechanics was the explanation of why the mean free path can be so much larger than the distance between neighboring atoms (of the order of 10^{-10} meter).

In good metals, n is the density of valence electrons, and is thus independent of temperature (except for very small temperature dependence due to the thermal expansion). Because the electrons follow the *Fermi-Dirac* distribution law, the typical velocity v is the velocity at the Fermi surface, which is large (usually about 10^6 meters/second), and independent of temperature. The temperature dependence of the conductivity is that of the relaxation time τ, and the approximate additivity of scattering reates due to different processes results in *Matthiessen's rule* which states that the resistivity is approximately the sum of a temperature dependent part R_T, due to scattering by lattice vibrations, and a part R_I proportional to the concentration of impurities and lattice defects. As the amplitude of the lattice vibrations increases with increasing temperature, the scattering effect and R_T increase. Above the Debye temperature θ (given by $k\theta = h\nu_m$, where k is the Boltzmann's constant and ν_m is the maximum frequency of the lattice vibrations), R_T is proportional to the absolute temperature. At low temperatures, R_T is proportional to the fifth power of the absolute temperature, and at all temperatures it is well approximated by the *Bloch-Gruneisen formula*. At very low temperatures (1 to 18°K), some metals become superconductors and all measurable resistance disappears.

Pure semiconductors at absolute zero temperature have no free electrons. As the temperature is increased, some electrons are excited to current-carrying states (in the *conduction band*). The states that are left unoccupied (in the *valence band*) are also free to carry current, and are called *free holes*. The concentrations of electrons and holes increase very rapidly with temperature, causing the resistivity to decrease. The carrier concentration may be expressed as

$$n = f(T)\exp(-\Delta E/2kT)$$

where T is the absolute temperature, ΔE is the energy gap between the valence and conduction bands, and $f(T)$ is a slowly varying function of temperature. Extra carriers may also be provided by impurity atoms, *donors* contributing electrons and *acceptors* trapping electrons and producing holes. Very small impurity concentrations may make very large changes in resistivity. If most of the free carriers come from impurities, the semiconductor is called *extrinsic*; it is n (negative)-type if electrons predominate, or p (positive)-type if holes predominate. If most of the carriers come from thermal excitation, the concentration of electrons and holes are about equal, and the material is called *intrinsic*. The rate of change with temperature of the mobility of a semiconductor is generally less than the rate of change of the carrier concentration. The mobility is often represented by $\mu = cT^r$, where r varies from about -2.2 to $+1.5$, depending upon the particular material, concentration, and type of defects. The conductivity of intrinsic material can also be expressed as

$$\sigma = g(T)\exp(-\Delta E/2kT)$$

where $g(T)$ is another slowly varying function of T. This equation is often used to analyze experimental data and find ΔE.

Insulators may be thought of as semiconductors with such large energy gaps that n is very small and the resistivity is very high. In addition, the ions in some insulating solids (such as the alkali halides) are free to move and carry a measurable current. The ions move by "hopping" into vacant lattice sites. As they must cross a potential energy barrier, the ionic mobility is proportional to an activation factor $\exp(-\epsilon/kT)$. The resistivities of such solids are of the order of 10^2 to 10^8 ohm-meters at room temperature, but are as low as 10^{-3} to 1 ohm-meter at elevated temperatures.

Another mechanism for electronic conductivity seems to operate in certain oxides and perhaps in some organic semiconductors. In this case, the electrons are localized and move by "hopping" in the same way as the ions in ionic solids.

Because the resistivity depends upon the state of crystalline order, it is used as a tool in studying phase changes, such as solid-liquid, order-disorder, magnetic, and crystal structure transitions. The resistivity of some materials is affected by pressure and mechanical strain. Many insulators and semiconductors exhibit the phenomenon of photoconductivity, in which the absorption of light produces free carriers and increases the conductivity.

Many conductors show the phenomenon of *magnetoresistance*, which is the increase in resistance when the conductor is placed in a magnetic field. (A very few materials exhibit negative magnetoresistance, which seems to be related to an inhomogeneous structure of the conductor.) The basic cause of the magnetoresistance is the Lorentz force, which causes the electrons to move in curved paths between collisions. Even for isotropic materials, the conductivity and resistivity must be taken as tensors in the presence of the magnetic field, called the *magnetoconductivity tensor* and the *magnetoresistivity tensor*. The off-diagonal elements of the magnetoresistivity tensor are related to the *Hall effect*. The on-diagonal elements for isotropic materials or for the magnetic field parallel to a crystal axis are usually called the longitudinal magnetoresistance (for the current parallel to the magnetic field) and transverse magnetoresistance (for the current perpendicular to the magnetic field). For small values of the magnetic field, the change in resistance is proportional to the square of the magnetic field strength. For large magnetic fields, the resistance may saturate (approach a constant value), continue to increase as the square of the magnetic field strength, or follow a more complex behavior. The magnetoresistance ratio (change in resistance divided by the zero-field resistance) reaches only a few percent for most materials, but at low temperatures for some materials, such as bismuth, approaches 10^6. In general, the transverse effect is larger than the longitudinal one; the effect is larger in high-mobility materials, and is largest for materials with more than one type of charge carrier. The magnetoresistance may be used in conjunction with the Hall effect to deduce the type, density, and mobility of charge carriers, and to obtain information about the Fermi surface.

Electrolytes. The conductance of an electrolyte has the same general definition as that given for any conductor. The same unit is used, the reciprocal ohm (mho). The term conductivity also has the same significance for electrolytes as for solid conductors, being conductance per unit length of path. Another term applied to electrolytes is *equivalent conductance*, which is the product of conductivity and the volume (in cubic centimeters) of solvent containing 1 gram-equivalent weight of solute at a specified concentration, measured when placed between electrodes 1 centimeter apart. The *molar conductance* is defined similarly, except that the weight of solute is 1 gram-molecular weight instead of 1 gram-equivalent weight. The *ionic conductance* is the amount contributed by each characteristic ion to the total equivalent conductance in infinite dilution. Thus, in the mathematical expression of the law of independent migration of ions:

$$\lambda_0 = \lambda_+ + \lambda_-$$

in which λ_+ and λ_- are the ionic conductances of cation and anion, respectively, and λ_0 is the total equivalent conductance of the electrolyte. The *conductance ratio* is the ratio of the equivalent conductance of a given ionic (electrolytic) solution to its equivalent conductance at infinite dilution.

See also **Electrolytic Conductivity and Resistivity Measurements;** and **Superconductivity.**

ELECTRICAL CONDUCTORS. See **Electrolytic Conductivity Measurements; Semiconductors; Superconductivity.**

ELECTRICAL INSTRUMENTS. The basic electrical analog instrument can be traced to Oersted's discovery (1819) of the relation between current and magnetism. Faraday (1821) learned that a current-carrying conductor would rotate in a magnetic field. Ampere (1821) demonstrated that a current in one conductor attracted or repelled another current-carrying conductor. Sturgeon (1836) wound the current-carrying wire into a coil and suspended it in a magnetic field. Kelvin (1867) placed a soft iron core in the center of the coil, shortening the air gap, increasing the sensitivity, and improving the scale characteristics of the device. D'Arsonval (1881) patented an instrument of this type. Weston (1888) discovered the factors required to produce a permanent magnet system, added soft iron pole pieces, devised current-carrying control springs and produced the first commercial double-pivoted permanent magnet moving coil instruments as such. From Oersted's and Kelvin's work also stem the principles on which polarized iron vane and electrodynamic type instruments evolved through the work of Ayrton and Perry (1881), Bruger (1886), Weston (1889) and many others.

Traceability

Electrical instruments are calibrated in terms of the prime standards of the quantity measured. These prime standards are established and maintained by various government standards organizations throughout the world—sometimes referred to as legal standards. The national caretaker of standards in the United States is the National Bureau of Standards. The method of tracing the accuracy of voltmeters, ammeters, and wattmeters to prime standards is outlined schematically in Fig. 1. Sophisticated standards and calibration services are also available from a limited number of private firms. Government laboratories generally calibrate only primary standards in selected areas; private firms generally must be relied upon to calibrate instruments, test equipment, or standards in other areas. A directory of standards laboratories is available from The National Institute of Standards and Technology (NIST), Physical Measurements and Services Program, Gaithersburg, Maryland 20899.

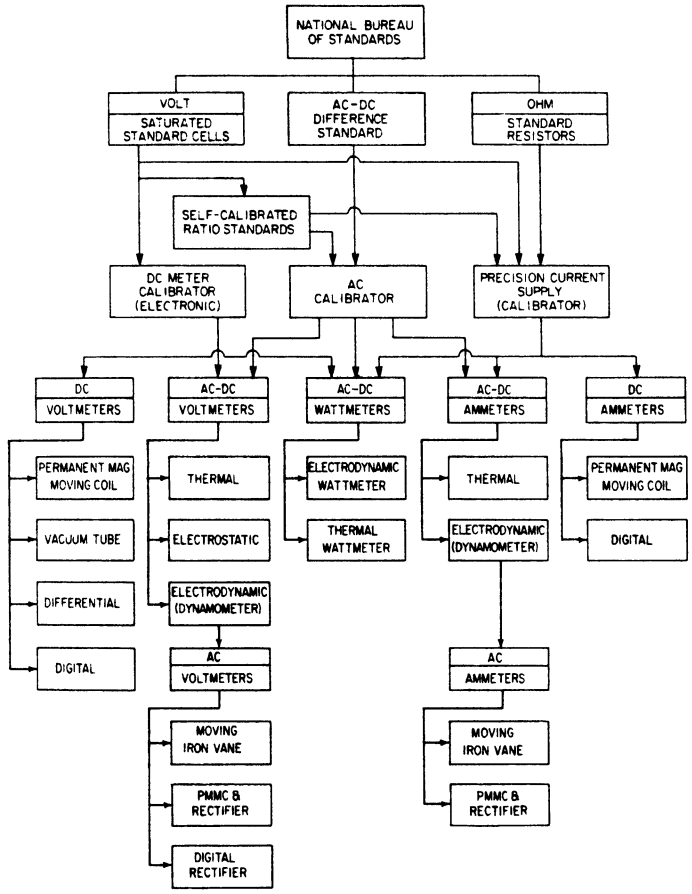


Fig. 1. Traceability of electrical instruments back to prime standards maintained by the National Institute of Standards and Technology and by similar standards organizations in other countries.

Manufacturers of standard cells and precision electrical instruments and systems usually maintain their own standards laboratories which contain local standards that are calibrated or recalibrated on a periodic scheduled basis by an official organization, such as NIST; secondary standards and the working standards used in various stages of manufacturing and assembly are, in turn, calibrated against these local standards.

Various types of ratio devices are used to extend the standard-cell voltage (approximately one volt) to cover many ranges of commercial instruments and in terms of current in amperes in addition to voltage through the use of standard resistors. Extension of voltage values is accomplished through the use of what is commonly referred to as a volt box. Effectively, this is a high-valued resistor, accurately tapped at exact proportions of its total value. If the total voltage is applied to the total resistance, then the value at the tapped points is low enough to compare accurately with the standard-cell voltage.

Extension of current values is accomplished through the use of very low-valued resistors; passing the unknown current through them results in voltage which, in turn, can be compared with the standard-cell voltage, and through the use of Ohm's law and the resistance value, the value of the current can be established. Such low-valued resistors are called *shunts* and standard units of this kind are maintained with much care. These shunts can be calibrated—as by the National Bureau of Standards.

The direct voltages and currents established at this level are used to calibrate electronic calibrators which, in turn, are used to calibrate dc voltmeters, ammeters, and ac-dc voltmeters, ammeters, and electrodynamic wattmeters. Also shown in Fig. 1 are ac-dc difference or transfer standards. These are high-accuracy instruments, such as the thermaland electrodynamic-type voltmeters, ammeters, and wattmeters which are sent to the NIST for calibration. In this case, NIST does not perform an accuracy check, but rather determines the difference between the ac calibration at various frequencies and that on direct current. These standards then are used for determining the performance of ac-dc instruments when used on alternating current.

Alternating-current voltmeters and ammeters usually are calibrated against electronic ac calibrators which have been calibrated against direct current standards by using thermal transfer standards.

Voltage Standards. The cadmium cell, developed by Dr. Edward Weston in 1893 and universally known as the Weston Standard Cell, was used as a world standard of emf from the time of its acceptance by the International Committee on Electrical Units and Standards in 1908 until the advent of Josephson voltage standards in the 1970s. Standard cells are still the prevalent primary voltage standard in industry. The unit of voltage at the International Bureau of Weights and Measures and the legal voltages of most industrial nations are now maintained through the ac Josephson effect. The Josephson effect is realized through weak connections (commonly through a dielectric tunnel barrier a few nanometers thick) between superconductors. Such connections are referred to as Josephson junctions. For a wide range of frequencies, the current-voltage characteristics of a Josephson junction that is irradiated with electromagnetic radiation of frequency f will exhibit zero-resistance parts (constant-voltage steps) at voltages given by

$$V = nhf/2e$$

where n is an integer, h is Planck's constant, and e is the magnitude of the charge of an electron. The units of voltage are defined through the equation above by assigning a value ot $2e/h$. For the United States legal volt, V_{NBS}, the assigned value is $2e/h = 483\ 593.42$ GHz/V_{NBS}; for the international volt, V_{69-BI}, the assigned value is $2e/h = 483\ 594.0$ GHz/V_{69-BI}. In general, the ac Josephson effect is used periodically to determine the emf of a group of standard cells that then embody the unit of voltage between determinations.

A standard cell is made in two types: (1) the normal cell containing a saturated cadmium sulfate solution, and (2) a type used as a working standard in which the solution is less than saturated above 4°C. The saturated cell is the basic standard for maintaining the value of the volt, and is used in this manner in all national laboratories. Its rather high temperature coefficient must always be taken into account.

The emf of the unsaturated cell is within 1.0188 and 1.0198 abs volts, the exact voltage of each cell being established by comparison with the normal or saturated cell. This cell is a useful working standard because of its negligible temperature coefficient.

Electrochemical cells of the type described have a relatively high internal resistance and thus should be used under zero load conditions. The cells should not be short-circuited because several weeks may be required for full recovery of a stable reference voltage, i.e., if the cell has not been permanently damaged.

Highly precise reference signals are available from electronic standards which employ temperature-compensated zener diodes. A wide variety of precision black box instruments, arranged for flexibility and convenience of use, is marketed.

Standard Resistors. The hermetically sealed, wire-wound one-ohm resistors designed by J. L. Thomas of then National Bureau of Standards are among the highest-quality resistance standards ever produced. The United States legal ohm, Ω_{NBS}, is the mean four-terminal resistance of a group of Thomas-type standard resistors immersed in oil at 25°C under a power dissipation of 0.01 watt. In the Thomas-type standard, the container is made of coaxial cylinders only slightly different in diameter with the space between the cylinders sealed. The resistance element (carefully annealed Manganin™ wire) is mounted in this space in good thermal contact with the smaller cylinder, which serves as the inner wall of the container.

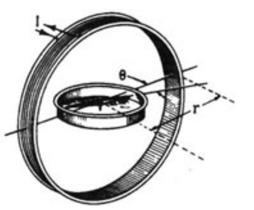
A subgroup of five Thomas one-ohm resistors, together with resistors of different values and/or different design, are used as working standards at NIST. These standards enable calibrations to be made over a wide range in resistance values ($10^{-4}-10^{15}$ ohms) and under conditions differing from those under which the legal unit is maintained. Examples of the latter are standard resistors especially designed for stable characteristics at very high currents.

Anatomy of Electrical Measuring Instruments. Although analog-type indicating instruments still are found on machines and in the laboratory, digitalization of circuitry and displays, permitting better readability and portability, created a sharp demand during the last few decades, and, as of the early 1990s, most catalogs will feature the digital variety of meters. It is very helpful in understanding electrical measurements to review briefly the chronology of electrical measurement concepts.

Polarized Iron-Vane Mechanism

The polarized iron-vane system was the first current measuring mechanism to be reduced to practice. Originally dependent upon the earth's magnetic field to provide the control torque, it was later made both more sensitive and more convenient in use by the use of a permanent magnet. In the form shown in Fig. 2(e), the mechanism came into use in large quantities as an automobile dashboard ammeter. Better suited to mass production in relatively crude form than to greater refinement, this mechanism has lost its position in the field of electrical instruments and has, instead, made a place for itself in the "indicator" field where precision is not demanded.

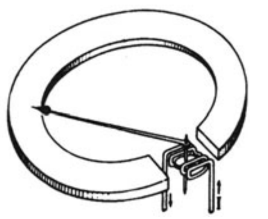
(a) The original electrical indicator—Oersted—1918



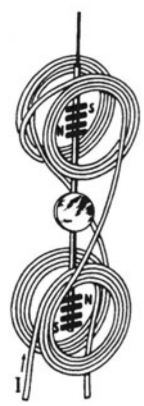
(b) Tangent Galvanometer.

$$I = \frac{10H}{2\pi N} \tan\theta \text{ amperes.}$$

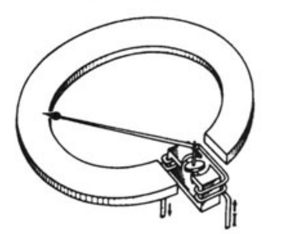
r = radius of coil.
H = horizontal component of earth's field in gauss.
N = number of turns



(c) Simple polarized iron vane mechanism. Pointer is driven by an iron vane governed by the resultant field of the permanent magnet and that of the current in the coil.



(d) Early type astatic galvanometer —Kelvin—1858. Used on early trans-Atlantic cable circuits. Note lower set of needles has reversed magnetic polarity from upper set, thus reducing the restoring force of the earth's field land increasing the sensitivity.



(e) Complete polarized iron vane mechanism. Soft iron core adds to sensitivity, requiring fewer turns of wire and improves scale characteristics. Weston prototype——Model 354 (now obsolete).

Fig. 2 Forms of polarized iron-vane mechanism.

Moving Iron-Vane Mechanism

In an early magnetic vane mechanism of the suction type the opposing force or restoring torque is provided by gravity instead of the presently used conventional spring. This method was widely used in older instruments. All gravity-controlled instruments had the major disadvantage of being subject to serious position errors.

With reference to Fig. 3, if two similar adjacent iron bars are similarly magnetized, a repelling force is developed between them which tends to move them apart. In the moving iron vane mechanism, this principle is used by fixing one bar in space and pivoting the second so that it will tend to rotate when the magnetizing current flows. A spring attached to the moving vane opposes the motion of the vane and permits the scale to be calibrated in terms of the current flowing.

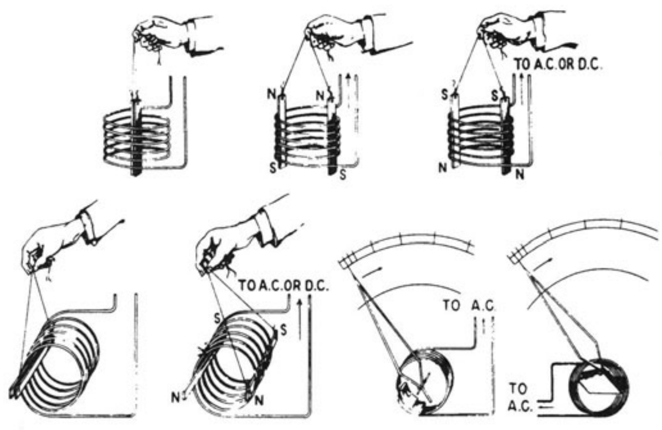


Fig. 3. Principle used in moving iron-vane mechanism.

In the concentric vane mechanism shown in Fig. 4, vanes slip laterally under repulsion. This design has only moderate sensitivity and has square-law scale characteristics. The short magnetic vanes result in small direct-current reversal and residual magnetism errors. With this mechanism, it is also possible to shape the vanes to secure special characteristics, thereby opening the scale where needed.

The radial vane mechanism, shown in Fig. 5, opens up like a book under repulsion. Of the polarized iron-vane mechanisms, this is the most sensitive, has the most linear scale, and requires better design and better magnetic vanes for good grades of instruments. An aluminum damping vane, attached to the shaft just below the pointer, rotates in a close-fitting chamber to bring the pointer to rest quickly.

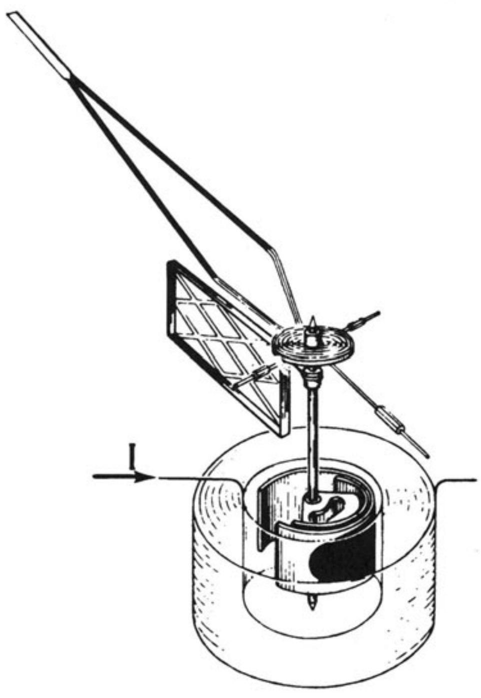


Fig. 4. Concentric vane mechanism.

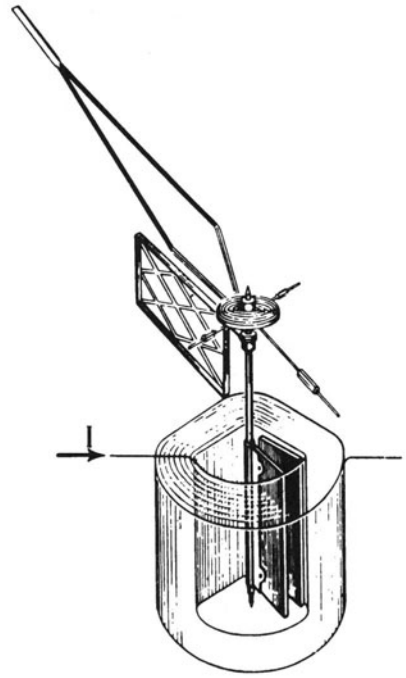


Fig. 5. Radial vane mechanism.

The Electrodynamic Mechanism

Electrodynamic mechanisms are the most fundamental of all of the indicating devices presently used. Form factor variations do not occur because of the complete absence of magnetic materials, such as iron, and the indications are of true RMS values. This mechanism is current sensitive—the pointer moves because of current flowing through turns of wire. It is the most versatile of all of the basic mechanisms since the single-coil movement can be used to indicate current, voltage, or power, ac or dc. Crossed-coil movements can be used for power factor, phase angle, frequency, and capacity measurements.

An early version of this basic principle was the current balance of Lord Kelvin (1883) and shown in Fig. 6.

Perhaps the most important use of this mechanism is as a transfer instrument between the basic standards of E, I, and R, all of which are defined for direct current only, to alternating current, in which form most of the power of the world is generated, sold, and used. The torque produced in the moving coil of an electrodynamic instrument is proportional to the product of the in-phase components of the currents in the field and moving coils; this mechanism measures products.

The electrodynamic mechanism used in wattmeters, voltmeters, and ammeters is shown in Fig. 7. With a longer pointer and other minor changes, the mechanism is used in a laboratory standard. With the further modification of crossed-moving coils, fundamentally the same mechanism is used for measurements of power factor, phase angle, and capacity. In this form or with crossed field coils, this mechanism measures a ratio by balancing two torques.

Damping vanes similar to those used in moving iron instruments (Figs. 4 and 5) rotate in close-fitting chambers and provide proper damping for the pointer motion. If two complete field coil systems are arranged one above the other, each including and acting upon its own moving coil, and if the two moving coils are attached to a common shaft which carries the pointer, the mechanism can be used to measure the total power in a poly-phase ac system.

The Electrostatic Mechanism

The electrostatic mechanism resembles a variable condenser. Of all the mechanisms used for electrical indications, it is the only one that

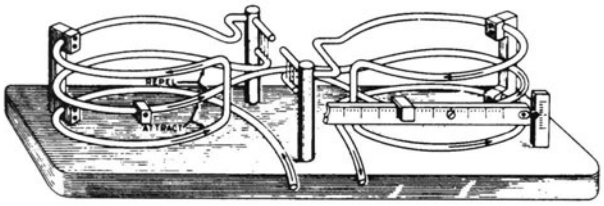


Fig. 6. Current balance of Lord Kelvin.

measures voltage directly rather than by the effect of current. The torque resulting from the attraction between fixed and moveable plates is a function of the voltage between the plates, the plate area and, inversely, the distance between the plates. For greater sensitivity, this distance must be reduced, clearances permitting, or the plate area (and thus the weight) must be increased. Greater weight, in turn, calls for still greater plate area if a sufficient torque is to be developed to overcome pivot friction in most industrial applications. This limits the use of the instrument to certain special applications, particularly in ac currents of relatively high voltage, where the current taken by other mechanisms would result in erroneous indications. A protective resistor is used in series with the instrument to limit current flow in the event of a short between plates.

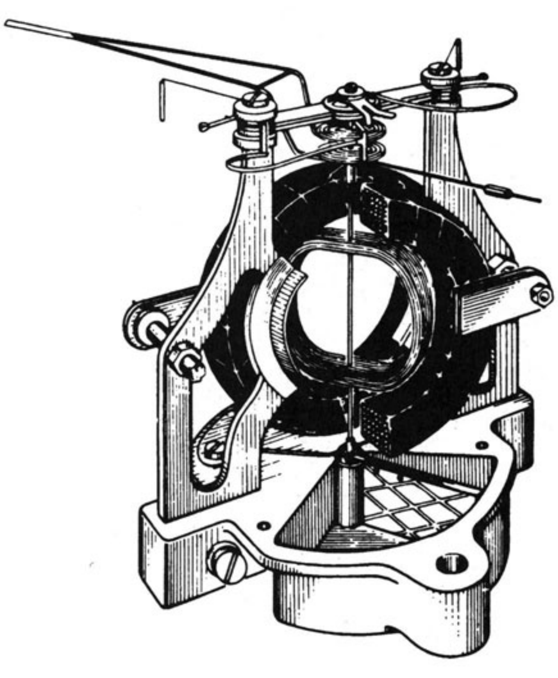


Fig. 7. Electrodynamic mechanism used in some voltmeters, ammeters, and wattmeters.

A gold leaf electroscope is shown in Fig. 8. Like charges on the ends of the leaf cause a separation of the ends. This is very sensitive as an indicator, but not characterized as an instrument per se. A form of the attracted-disk electroscope as devised by Sir William Snow Harris (1834) is shown in Fig. 9. Guard ring around the disk served to prevent nonuniformity in the electrostatic lines of force. A very large example of the attracted-disk electrometer mounted in a shielded cage and using quartz pillar supports is used at the National Bureau of Standards for voltage standardization up to 300,000 volts. Using this high voltage electrometer, the ratio of transformation of high voltage potential transformers has been checked for the first time by an independent method.

Fig. 8. Gold leaf electroscope.

Fig. 9. Attracted-disk electroscope.

Figure 10 illustrates the principle of an early electrostatic mechanism devised by Lord Kelvin in 1887. Moving and fixed plates attract each other causing the moving plate to rotate against the torque of a control spring (not shown). Position of balance thus becomes a measure of the potential applied. In industrial form (Fig. 11), such instruments are made with multiple sets of plates in ranges from 150 to 3,500 volts.

Permanent-Magnet Moving-Coil Mechanism

The design shown in Fig. 12 offers the largest magnet in a given space and is used where maximum flux in the air gap is required in order to provide an instrument of lowest possible power consumption.

With the advent of improved magnetic materials, it became practical to design a magnetic system in which the magnet serves as a core. Such magnets operate at their highest energy product with minimum lengths, thereby making the core magnet mechanism a practical reality. These mechanisms have the advantage of being relatively resistant to external magnetic fields, therefore allowing for the elimination of magnetic shunting effects in a panel or the need for magnetic shielding in the form of iron cases. See Fig. 13.

The advantage of the self-shielded feature makes the core magnet mechanism particularly useful for aircraft and aerospace applications, especially where a multiplicity of mechanisms must be mounted in close proximity to each other as exemplified by cross-pointer indicators wherein as many as five mechanisms are mounted in one case to form a unified display.

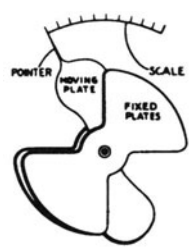


Fig. 10. Early electrostatic mechanism devised by Lord Kelvin.

Fig. 11. Moving-and-fixed plate-type electrostatic instrument.

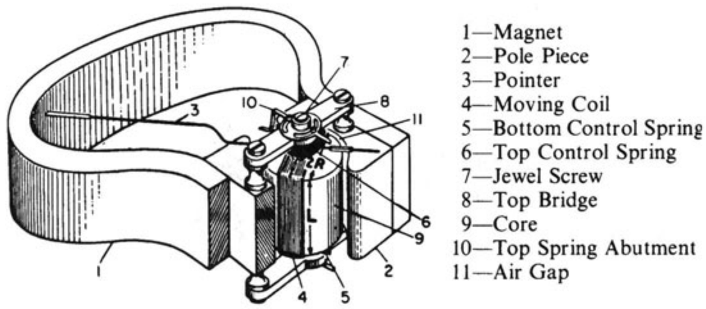


Fig 12. Permanent-magnet moving-coil mechanism with external magnet.

End-Pivoted or Off-Center Coil Type. The end-pivoted permanent-magnet moving-coil mechanism is a variation of the more common center-pivoted type. In this arrangement, the coil rotates in a single air gap, allowing a full-scale deflection as high as 270°. The concept is not new. Meters of this type were made as far back as 1900, but with low magnetic flux and poor performance. An improved instrument has a scale deflection of 250°. The magnet extends into the soft iron core, thereby minimizing leakage flux and achieving a high flux density.

The off-center coil type is also used extensively in edgewise panel and aircraft meters. The deflection is usually limited to 60° or less with the magnetic flux concentrated over the smaller angle.

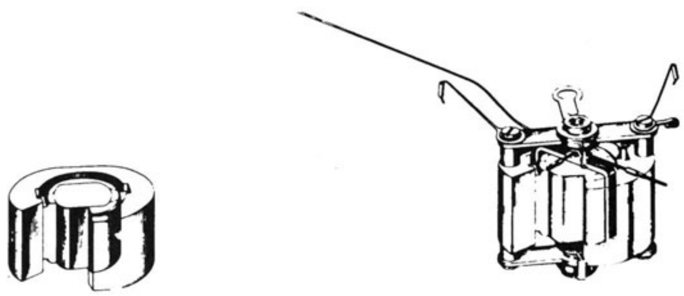


Fig. 13. Core magnet and moving-coil mechanism.

Taut-Band Instruments. The suspension type mechanism has been known for years. Until recently, the device was utilized for laboratory equipment where high sensitivities were required and available torques were extremely low so that it was desirable to eliminate even the low friction of pivots and jewels. The device had to be used in the upright position since the sag in the very low torque ligaments caused the moving system to come in contact with stationary members of the mechanism in any other position. The taut band instrument enables one to obtain the advantage of the absence of friction of ribbon suspension while also eliminating the sag by placing the ribbons under sufficient tension. This tension is provided by a tension spring so that the instrument can be used in any position. Generally speaking, taut band suspension instruments can be made with higher sensitivities than those using pivots and jewels and can be utilized in virtually every application which is presently served by pivoted instruments. See Fig. 14.

A permanent-magnet moving-coil mechanism is not insensitive to temperature by itself but may be made so by the appropriate use of proper series and shunt resistors of copper and manganin. Magnets and springs decrease in strength and copper increases in resistance with increase in temperature. The changes in the magnet and the copper tend to make the pointer read low on fixed voltage impressed, while the spring change tends to cause the pointer to read high. The effects are not identical, however, with the result that an uncompensated mechanism tends to read low by approximately 0.2% per degree centigrade. For purposes of specification an instrument is considered compensated when the change in accuracy due to 10 degrees change in temperature is not more than one-fourth of the total allowable error.

Wattmeters, thermal watt converters, watthour meters, power factor meters, var meters, and instrument transformers are described in the entry on **Electric Power and Energy Measurement.**

Check Alphabetical Index for other electrical instruments described in this Encyclopedia.

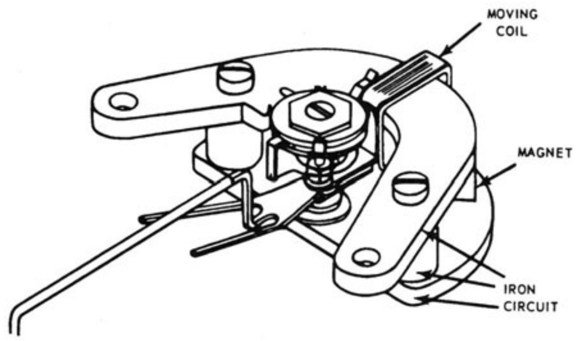


Fig. 14. Taut band suspension.

ELECTRIC CARS. The tempo of research and development engineering for mass production of battery-operated highway vehicles has undergone a marked increase during the past decade. Experts in the field now predict that substantial numbers of electric cars will be registered as early as 1998 and that these cars will constitute 5% of all cars on the road in the United States and of other advanced countries of the world by the early years of the 21st century. Electric cars will reduce air pollution, particularly troublesome in large metropolitan areas, and ultimately will dramtically reduce the petroleum consumption demanded by internal combustion engines.[1] Although trial runs of car and truck fleets have been conducted for several years, to date these have been essentially regarded by the general public as curiosities. Altough much further progress in battery and power train design is required to build and market electric vehicles to attract buyers, some professional people in the field believe that the mid-1990s will witness a marked increase in public interest and certainly of awareness.

A Matter of Energy Balance. A key objective in designing the power package for an electrically powered vehicle has been that of finding a *high-density* power source, coupled with the objective of creating *zero emissions* into the atmosphere. The energy density problem is well illustrated by the case of using lead–acid batteries. A 20-gallon (75.7-liter) tank of gasoline can provide about 2.4 million Btus (0.6 million Calories). Lead-acid storage batteries, weighing about the same as that quantity of gasoline, can provide only about 7700 Btus (1940 Calories), or about $2\frac{1}{4}$ kilowatt-hours. The ratio of energy in a tank of gasoline to the energy in the same weight of conventional storage batteries (before they are fully discharged) is more than 300 to 1. However, an automobile engine can convert only about 20% of the energy in the gasoline into driving power, while the electric motor can produce motive effort from nearly all of the electricity delivered by a battery. This reduced the margin to 60 to 1, still a very marked disadvantage for lead-acid batteries as compared with the internal combustion engine.

Prior to briefly describing current battery research, it is in order here to review the lead-acid battery situation further—because there are par-

[1]This observation immediately poses the question, "Since fuel is required to create the electricity that will charge the batteries of electric cars, will this not simply increase the petroleum consumption of the electric generating plants?" There will be increased need for electric power, but power plants do not depend heavily (6–7%) on petroleum to fuel their boilers, whereas the internal combustion engines of autos consume some 63% of total petroleum in the United States.

allels even though, with the more promising batteries, the situation becomes much more favorable for the battery power source. If it is first assumed that lead-acid batteries are used to power passenger cars, experience indicates that an electric car in city traffic at speeds up to 35 miles (56 kilometers) per hour uses about $\frac{1}{4}$ kilowatt-hour per 10 miles (16 kilometers) traveled. If it is further assumed that the car weighs 3000 pounds (1361 kilograms), half of that weight must be allowed for the battery. The very best lead-acid batteries currently available can deliver 15 watt-hours per pound if discharged over 20 hours. Thus, the battery can deliver, at the maximum, 22.5 kilowatt-hours. However, the battery will be discharged over a 1- to 2-hour period and thus its capacity will be about halved. Thus, the useful available output at most will be 11 kilowatt-hours from a 1500-pound (680-kilogram) battery pack. Inasmuch as the car is assumed to weigh $1\frac{1}{2}$ tons (1361 kilograms), thereby using 0.375 kilowatt-hour per mile, the car will have a maximum range of 30 miles (48 kilometers) at 35 miles (56 kilometers) per hour with a fully charged battery. If, instead of 35 miles (56 kilometers) per hour, the operator elects to travel at 55 miles (88 kilometers) per hour, the range would be cut in half, or approximately 12 to 15 miles (19 to 24 kilometers) per charge.

The foregoing model depends upon certain assumptions that, through the use of available lighter materials of construction and improvements in motor and drive systems now underway, can improve the case of the battery somewhat. But, needless to say, the internal combustion engine trips the energy balance by a wide margin. Thus, in connection with new battery designs, energy density is a major design target.

Battery Types. Batteries with several differing components for their cells currently are under intense study. They range in the degree of perfection attained to date.

Lead–Acid Battery. It is of interest to note that General Motors (GM) selected a pack of 32 lead–acid batteries for the power supply of the Impact car announced in late 1991. It is claimed that this car, equipped with two 57-horsepower alternating current motors can achieve a top speed of 100 mph (161 kph), but is equipped with a governor to limit its speed to 75 mph (121 kph). The vehicle is claimed to accelerate from 0 to 60 mph (97 kph) in 8 seconds. The two motors, operating together, develop 114 HP at 6600 rpm. A range of 120 mi (129 km) is estimated. Details on lead–acid battery are given in the article on **Battery.**

Nickel-Iron Battery. Although the Ni–Fe battery dates back to Edison, who first constructed it, the battery has not been mass produced and, consequently for use in electric autos, much more manufacturing experience will be required. The battery was used in U.S. Navy submarines for a number of years prior to nuclear power. One negative factor is that it produces hydrogen gas, which is explosive at certain air-gas ratios. Engineers who are inclined toward the Ni–Fe battery are impressed by its ruggedness, and project that the battery would last for the whole lifespan of a vehicle. It is reported that Chrysler, in cooperation with EPRI (Electric Power Research Institute), is designing a van to be powered by 30 such batteries having a total weight of 1800 lb (816 kg). The Ni–Fe battery features 50% more energy density than its lead–acid counterpart.

Nickel–Cadmium Battery. The Ni Cd battery has enjoyed some favor among European car manufacturers, such as Volkswagen and Mercedes-Benz in Germany. Mercedes is testing six Ni Cd powered 190 models. BMW has built ten battery-powered model 320s.

Lihium Alloy–Metal Sulfide Battery. This battery, a cross section of which is shown in Fig. 1, is based on a relatively simple and safe technology that has made it attractive in the past for use in military equipment and other products for which performance and reliability are critical. The battery has a specific energy capability of 100 watthours/kilogram. This represents a tripling of the range estimated for vehicles that use improved lead–acid types.

Sodium–Sulfur Battery. A Ford Aerostar has been converted to battery power through the use of sodium–sulfur batteries that provide about three times the range of lead–acid batteries of the same weight. One negative factor is that these batteries must be maintained at very high temperature, estimated at about 600°F (316°C) for maximum efficiency. It is claimed that normal driving and charging activities keep the batteries at that temperature. An additional negative is their inoperability after about three days of no use, thus creating problems in connection with long-term parking. Designers are attempting to overcome permanent chemical changes that may occur when the batteries are cold. It has been reported that 100 sodium-sulfur equipped European Escort vans, with a top speed of 70 mph (113 kph), will be tested soon.

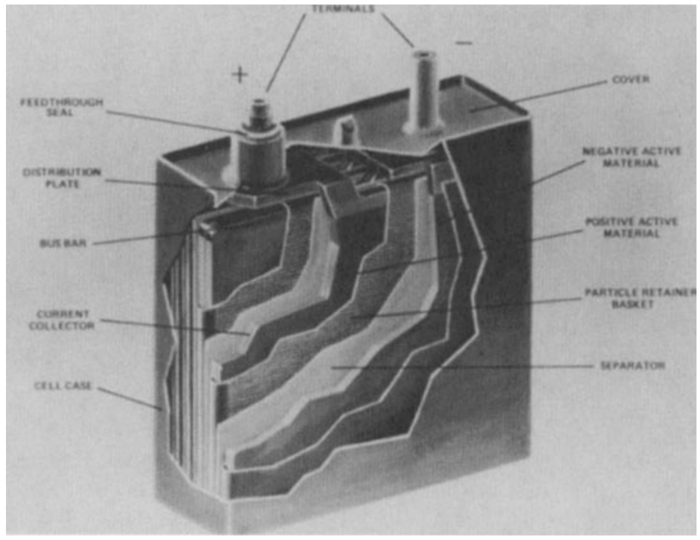


Fig. 1. Lithium alloy/metal sulfide cell in late stage of development and testing in the late 1980s. This is one of several new battery designs considered to have excellent potential for use in the electrical vehicles of the future. (*Electric Power Research Institute.*)

Comparison of Battery Performance. A rough comparison of performance estimates is given as follows:

Lead–Acid	Up to 1000 charges	250 mi	(400 km) range
Nickel–Iron	1100	300	480
Nickel–Cadmium	2000	325	525
Lithium–Aluminum/ Iron sulfide	600	470	755
Sodium–Sulfur	600	500	800
Lithium–Aluminum/ Iron disulfide	600	680	1095
Lithium polymer	700	740	1190

Power Trains. Much research is being directed toward improving the power trains of electric vehicles. These divide between the use of alternating current and direct current motors. One AC program involves the use of an advanced power transistor and integrated circuit technology. It has been claimed by some researchers to be superior to DC systems and competitive with other AC systems in cost, weight, and performance. In an electric van fleet tested, the AC power train, when compared with superior DC power trains, showed a decrease in initial costs by about 5% and increased driving range by about 10%.

Battery-charging Methodologies. Several schemes have been suggested pertaining to how the electric car user will be able to retain a sufficient charge in the batteries to assure reasonable usage of the vehicle. The charging problem related to city driving, short-distance commuting, and the like is much simpler than for traveling long distances. For the former, it has been proposed that the user would maintain a charging facility in the residence and recharge the batteries during periods of low, general power consumption, a desirable off-peak situation for the utilities. Or, during the day, charging facilities would be available at the office building or factory, wherever the user works during the day. For long trips, it has been suggested that there would be charging stations along the highway, probably facilities added on to the normal service station that dispenses gasoline and diesel fuel. The other major battery recharging problem is designing a perfectly safe system that can be used in all kinds of weather, and, of course, a system that would charge the batteries quickly.

Not much has been said recently, but another system would exchange a fully charged battery for one that needs charging. This approach would require an additional design problem for the car manufacturer.

General Motors has announced a proprietary "inductive coupler" charging unit that is claimed to be safe and economic for "fueling" electric vehicles. With the Hughes-developed system, it is claimed that drivers will be able to recharge their vehicles simply by inserting a 5-inch (12.7-cm) round, plastic-covered paddle into a slot in the car. The system "transfers electricity through a magnetic field rather than a direct metal plug–to–metal socket connection. A spokesman for the firm has stated that the car's charging system can be accomplished safely and without fear of shock in rain and very moist conditions.

Hybrid Systems. A typical hybrid vehicle is equipped with a small generator that is powered by a small internal combustion engine. This engine-generator, which provides the vehicle's average power requirement, lengthens the mileage over which the battery's energy is expended. Coupled with any of the newer battery designs, the engine-generator concept could provide a cost-effective way to increase vehicle range when long-distance travel is occasionally needed. Routine, shorter-range travel would be powered by the propulsion battery alone.

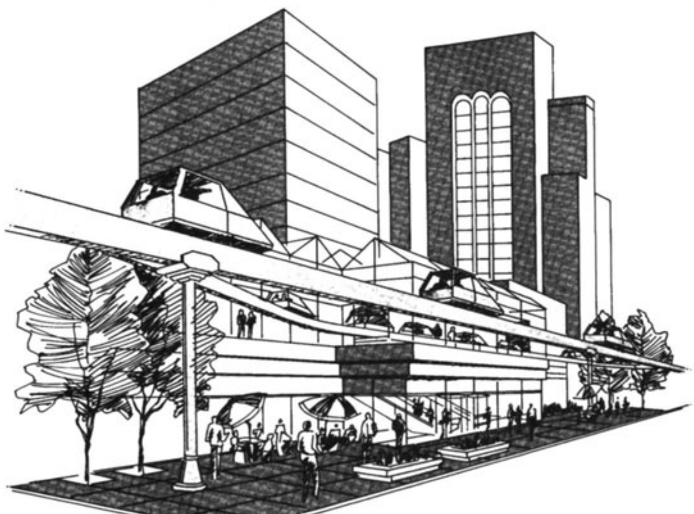


Fig. 2. Personal rapid transit system designed as an alternative to private automobiles. A test segment of the system is scheduled for testing in a midwestern city in the United States during the late 1980s. (*Electric Power Research Institute.*)

Incentives and Projects. Although car manufacturers have displayed numerous earnest efforts to develop an electrically powered vehicle that will attract buyers, an underlying incentive is legislative. This is exemplified by a current law on the books in California which requires any automaker selling more than 5000 cars annually, commencing in 1988, to include within the total at least 2% of "no emissions" vehicles. Progress has been made with solar-powered vehicles, but the technology would be far short in targeting for such an early deadline. Hydrogen-fueled systems (which emit water vapor only) will require development extending well into the next century. Thus, the battery-operated vehicle is the only viable candidate in view.

Most of the research in this field has been conducted by the large automakers in the United States, Japan, and Europe, battery manufacturers, and universities working on grants. Most vehicle makers have supplemented their central research in recent years with impressive electronics capabilities which include programs for developing electric vehicles. Because batteries require electric power for their maintenance, some of the electric utilities also have been active in the electric vehicle area. For example, EPRI (Electric Power Research Institute), an organization sponsored by over a hundred electric utilities in the United States, has included in its recent annual research budgets over $1500 million for research on batteries and their infrastructure. The goal: "In cooperation with the auto and utility industries, focus on developing batteries and infrastructure that could result in the development of more than 2 million electric vehicles by 2002, with an intermediate step that could deploy over 200,000 electric vehicles by 1998." If fully developed, this will result in electric power sales increases of 10 billion kilowatt hours/year; a carbon dioxide (CO_2) reduction of 220,000 tons/year; a nitrogen oxides (NO_x) reduction of 37,000 tons/year; a carbon monoxide (CO) reduction of 8 million tons/year; and a volatile organic compounds (VOC) reduction of 22,000 tons/year.

Fig. 3. The Electrified Roadway concept is scheduled for testing in a west coast city prior to 1990. Shown here is a prototype bus that will operate over a one-mile segment of electrified road. The propulsion battery is designed to be recharged as the bus travels over the electrified road segment. This will enable the bus to operate all day on both electrified and non-electrified segments of the proposed three-mile route. (*Electric Power Research Institute.*)

Long-Term Potential. Among many other concepts for improved urban and suburban travel of the future is the Electrified Roadway. These roadways would be suitable for use with electric cars, trucks, and buses. An electric cable buried beneath the road surface would provide electricity through inductive coupling to a pickup on the vehicle. This would simultaneously propel the vehicle and charge its battery. Automated vehicle control schemes could also be integrated into the electrified roadway system. See Figures 2 and 3.

Historical Perspective. Records show that 1575 electric automobiles were built in the United States in the year 1900, while only 939

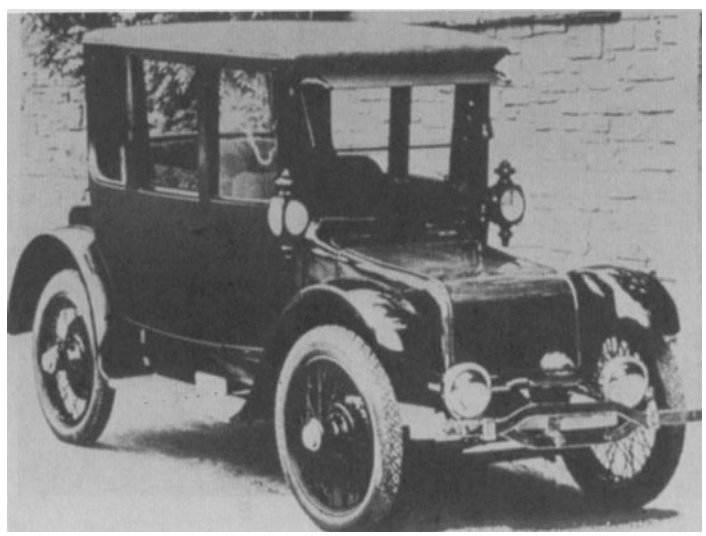


Fig. 4. The concept of the electric automobile dates back many years. Vehicle shown here was popular during the early part of the 20th Century. Electrical vehicles offer numerous advantages in terms of economy, reduced air pollution, lower cost of maintenance, and longer useful life—provided that battery and powertrain improvements presently underway can be achieved. Among other programs, as of the late 1980s, a fleet of GM Griffon electric vans is undergoing tests in selected regions of North America. (*Electric Power Research Institute.*)

cars equipped with gasoline engines were manufactured. The feasibility of the electric car was established well over eight decades ago. Electric automobiles were particularly popular with women during the early part of this century because of the difficulty in starting piston engines with a crank (prior to the self-starter). See Fig. 4. Coupled with gross improvements in gasoline-powered engines, the self-starter eclipsed the future of the electric car for many years. By 1930, the electric automobile was largely regarded as a museum piece. The principle of electric propulsion has been applied over the years and to the present, however, in the form of hundreds of thousands of electric trucks, fork lifts, etc. used in numerous materials handling operations. Because these off-road vehicles operate at low speeds and frequently only during one shift out of 24 hours, the standard lead-acid storage battery was adequate.

ELECTRIC CIRCUITS. Electric charges, at rest and/or in motion, are the fundamental sources of electric and magnetic fields. Broadly speaking, there are three more or less distinct situations involving the spatial distribution of the fields produced by distributions of charge and current. Many important instances arise where these fields are distributed throughout a region of space of vast extent and differ accordingly at separated points in the region. The fields produced by a radio antenna come under this category. On the other hand, important applications are made of devices in which the electric and magnetic fields are confined to a much more limited region of space although the fields still undergo a significant variation in magnitude from point to point throughout the region. A transmission line connecting a television antenna on a roof to its associated receiver is an example of this type of situation. Here the fields are confined to the immediate vicinity of the line conductors but the fields vary significantly as one moves along the line. Finally, however, there are applications utilizing pieces of apparatus in which the electric and magnetic fields are confined to regions of space which are so restricted in spatial extent that one may speak of an electric or magnetic field as having an essentially constant value in the immediate region of the device which is very much greater than at all other points of space. As an example it is noted that if a large number of turns is wound on a small circular core to form an inductance through which a direct current passes, a magnetic field will be created which is concentrated in the immediate vicinity of the coil and vanishes rapidly as one recedes from this location.

It is found, however, that when one deals with changing currents (charges in motion with varying velocity), the spatial dimension alone is not an adequate measure to establish which of the three situations suggested above is involved in a particular application. For sinusoidally varying currents one can define a quantity known as the wavelength which in free space is numerically equal to the velocity of light divided by the frequency of the sinusoidal variation. If spatial extent is measured in wavelengths, then an assemblage of apparatus which is confined to a region which is less than about one tenth of a wavelength in greatest dimension results in a good approximation to the third situation considered above. This possibility, one in which the electric fields and the magnetic fields may be considered to be concentrated in individual pieces of apparatus, is the domain of electric circuits. An electric circuit may be defined as a characterization of an electrical system in terms of the integrated effects of the electric and magnetic fields present in the system. The characterization is an approximation to the actual field problem in which one replaces the actual system by elements having resistance, capacitance, and inductance and by sources of electric potential and electric current. Systems in which the approximation cited is permissible are sometimes called "lumped constant" circuits. Antennas and transmission lines, on the other hand, are often referred to as "distributed constant" circuits. The distinction between these two designations comes from the spatial variation of the electric and magnetic fields as outlined above.

As may be inferred from their definitions, the parameters of resistance, capacitance, and inductance are not necessarily independent of the currents and voltages impressed upon the elements of an electrical system. Whether the resistance of an element is a function of the current through it or not, the relation $e = Ri$ (R is resistance in ohms) is still valid for the connection between the voltage across an element and the current through it. On the other hand, corresponding relations for the coil and the capacitor become more involved if their parameters are dependent on current or voltage. For constant parameters of inductance (L in henrys) and capacitance (C in farads), the voltages and currents in pure inductive and capacitive components assume the form

$$e = L\frac{di}{dt} \quad \text{or} \quad i = \frac{1}{L}\int e\,dt$$

and

$$i = C\frac{de}{dt} \quad or \quad e = \frac{1}{C}\int i\,dt$$

respectively. If the parameters L and C vary with impressed voltage and current, the above relations are not valid and recourse must be made to the basic law of electromagnetic induction when coils are involved and to the expression for electric current as the rate of change of charge (derivative of charge with respect to time) for circuits containing capacitors.

Whether elements with constant or variable parameters are involved, one may employ Kirchhoff's Laws of Networks to formulate equations representing conditions of equilibrium between the applied voltages and/or currents and the quantities that result. See **Kirchhoff Laws of Networks.** The equilibrium equations may be formulated on the basis of voltages as independent variables and currents as dependent quantities, the system of equations being known as the mesh equations for the circuit. Alternatively, they may be written on the nodal basis, where the sources are current generators and the dependent quantities are nodal voltage differences with respect to a reference node. As an example of the process involved, consider the simple series circuit shown in Fig. 1 and the simple parallel circuit shown in Fig. 2. The single mesh equation characterizing the series circuit and the single nodal equation describing equilibrium in the parallel circuit are, respectively

$$E(t) = Ri(t) + L\frac{di(t)}{dt} + \frac{1}{C}\int i(t)\,dt$$

and

$$I(t) = \frac{1}{R}\,e(t) + \frac{1}{L}\int e(t)\,dt + C\frac{de(t)}{dt}$$

The mesh and nodal equations for more complicated circuits have a form similar to these equations with added dependent variable terms and, in general, additional independent variable terms. The equations may be solved by various mathematical means to yield the desired unknown quantities, currents for mesh equations and voltages for nodal equations.

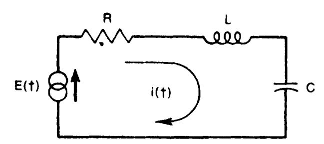


Fig. 1. Simple series circuit.

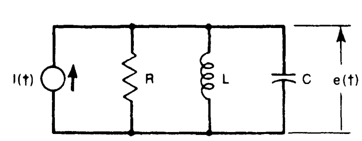


Fig. 2. Simple parallel circuit.

If the applied voltage sources are batteries or alternative sources of electric potential which are constant with time, the solution to the network equations described above will consist of constant terms (including zero as a special case) plus terms which decay exponentially with time. The latter terms are called the transient terms or the transient solution. They correspond to the solution of the homogeneous differential equations. The constant terms represent the steady state solution of the problem. These particular solutions may be obtained by the means that are considered in the section on Direct-Current Circuits where only networks of pure resistances are involved. (In the computation of the steady state response for circuits with constant potential applied, inductors may be replaced by elements of zero resistance and capacitors by resistors of infinite resistance.)

ELECTRIC CLOCK. A clock that employs electric current. Most automobile clocks use current from a battery to automatically rewind a mainspring. In other electric clocks, the mainspring is usually eliminated—the battery current driving the balance wheel through electromechanical contacts. An early line-current system used an auxiliary mainspring that was kept fully wound by the current for use should the line power fail. Later, a synchronous motor controlled by alternating current replaced the balance-wheel system in electric plug-in clocks.

Other types of electric clocks represent developments of the pendulum. See also **Pendulum clock.** An electromechanically-driven pendulum system, developed in 1843 by Alexander Bain in Scotland, was probably the first electric clock. At a considerably later date, the so-called Western Union clock used two 1.5-volt cells to wind a mainspring that maintained a pendulum while accepting an hourly time-pulse signal by telegraph line to synchronize its time display on the hour with a remote master clock system. The time-pulse signal, in effect, corrected for accumulated errors from all sources.

The later consumer cordless electric clocks use a 1.5-volt dry cell to drive the balance wheel electromagnetically, by means of an electronic circuit. By utilizing a transistorized electronic circuit as a switch to drive the balance wheel equipped with a tiny magnet, these clocks eliminate electromechanical contacts. In another clock, the frequency standard is an electromagnetically-driven 360-Hz tuning fork, powered by a single aspirin-size 1.3-volt mercury oxide cell. An electronic control circuit is required to drive the tuning fork. The accuracy of this clock averages plus-or-minus two seconds a day. If three tuning fork systems are electrically interconnected, as in a marine navigation clock, the daily deviation in rate drops to less than one second. See also **Tuning Fork.**

In the United States, plug-in electric clocks usually use self-starting shaded-pole synchronous motors, controlled by the 60-cycle alternating line current. Their accuracy, therefore, is dependent upon the line frequency. Throughout the United States, however, line frequency is normally both properly stabilized and corrected several times daily, eliminating accumulated error and permitting an accuracy of up to plus-or-minus four seconds. Where line frequency is not corrected at least once a day, the resulting error displayed by plug-in electric clocks can be far greater. In many locations elsewhere in the world, where line voltage is not normally stabilized, plug-in clocks are less reliable. Where powerline shutoffs are routine or frequent, plug-in clocks can represent a major inconvenience.

During the last few decades, the cost of manufacturing timing mechanisms has been reduced drastically. This has been coupled with the availability of very low cost crystal oscillators. A majority of table-top clocks today feature digital readouts, plus additional features for alarming. Even though these instruments do not depend upon line frequency, some models require power line current. Others utilize battery power. For those instruments, depending on power line current, power interruptions are a very serious source of inconvenience because loss of power for even a fraction of a second will require resetting. Most instruments are equipped to "blink" to indicate that the line source has been interrupted and that resetting is needed. A combination of line and battery power can eliminate the foregoing inconvenience when there is a power line interruption. Table-top electric clocks commonly are offered as a clock-radio combination.

ELECTRIC EEL. See **Gymnotid Eels.**

ELECTRIC FEEDBACK. In magnetic amplifier terminology, feedback through an electrically conductive network, as differentiated from feedback produced by currents in windings having coupling to the control windings (magnetic feedback).

ELECTRIC FIELD STRENGTH. The magnitude of the electric field vector. This term is sometimes called the electric field intensity, but such use of the word intensity is deprecated in favor of field strength, since intensity connotes power in optics and radiation.

ELECTRIC FIELD VECTOR. At a point in an electric field, the force on a stationary positive charge per unit charge. Under conditions in which the ratio of force to charge is not constant, the field vector is defined as the limit of the ratio as the change approaches zero. This may be measured in newtons per coulomb, in volts per meter or in corresponding units in systems other than the mksa system.

ELECTRIC FLUX DENSITY. At a point, the vector whose magnitude is equal to the charge per unit area which would appear on one face of a thin metal plate introduced in the electric field at the point and so oriented that this charge is a maximum. The vector is normal to the plate from the negative to the positive face. The term electric displacement density or electric displacement is also in use for this term. In an isotropic medium of permittivity ϵ, the flux density is $\mathbf{D} = \epsilon\mathbf{E}$ is rationalized mks units, where $\mathbf{E}$ is the electric field vector.

ELECTRIC INDUCTION. See **Induction (Electric/Magnetic).**

ELECTRICITY. An isolated atom consists of a small nucleus, itself composed of protons and neutrons, surrounded by a cloud of electrons. The proton and the electron are the ultimate stable particles of electricity. Their charges are equal and opposite, the proton being regarded, by convention, as positive. A normal atom with its full complement of electrons is thus uncharged.

Electrostatic phenomena arise when bodies (or parts of bodies) have an excess of electrons or protons, a state usually produced by transferring electrons, e.g., by means of a battery or by rubbing two dissimilar materials together. Between two positively charged bodies (or two negatively charged ones) there is a repulsive force; between positive and negatively charged ones, an attractive force.

If two bodies at different potentials are connected by a conductor, such as a metal wire in which there are free electrons, the electrons in the wire drift under the influence of the electric field. Such movement of electrical charges gives rise to further phenomena and we speak of an electric current. The current may be one of electrons only, as in a metal, in semiconductors, or in electron tubes. Or the current may be of positive nuclei, as in an isotope separator; or of both positive and negative charges, as in the conduction by ions (atoms that have gained or lost an electron) in liquids or in gaseous electrical discharges.

Electrons flowing in the positive direction give rise, by our convention of signs, to a negative current. In a metal, semiconductor, or conducting liquid, the velocity at which the electrons or ions drift is quite slow, less than one centimeter per second even for current densities in a metal as high as 10^4 amperes per square centimeter. In vacuum devices, such as cathode-ray tubes, the speed of the electrons approaches that of light. If a wire joins two electrostatic charges, the current persists for a short time only, but it may be maintained by means of some source of energy, such as a battery, a generator, a thermocouple, or a solar photoelectric cell.

When a current flows under the influence of a potential difference, the moving charges—electrons in metals, ions in solution—are impeded by collision with the atoms in the conducting metal or liquid. The charges give up to the atoms the energy they acquired as they moved in the electric field, and electrical power is converted into other forms—for instance, into heat in the case of a metal wire. Electric fields are also produced by time-changing magnetic fields, and this principle is extensively exploited, as is motional electromotive force (emf) to generate electric power. Electromotive force and voltage drop are usually regarded as synonymous. When an emf is impressed on a closed metallic circuit, current results.

The electrification of amber by rubbing with wool or fur was observed many centuries ago. Not until the work of Volta, late in the 18th century, was electricity recognized through any but electrostatic phenomena, and investigations on the properties and applications of electric currents were among the most brilliant features of nineteenth-century physics. Even in the 1890s physicists were still asking, "What is electricity?" It had then long been known that an appropriate application of energy will separate electricity into two components, designated as positive and negative; that bodies charged with these components attract each other; and that the energy of separation is yielded upon the reunion of the two components. It remained for J.J. Thomson to recognize the electron, and for the subsequent analysis of atomic structure to identify the proton, and to explain their relations.

Hundreds of entries in this volume deal with electrical fundamentals, equipment, and applications.

ELECTRIC LENGTH. The physical length of a transmission line or its equivalent, corrected for any inhomogeneities that may affect the speed of propagation, and expressed in wavelengths, radians, or degrees.

ELECTRIC MOTOR. See **Motor (Electric).**

ELECTRIC POTENTIAL. The electric potential at a point is the work done in moving a unit positive charge from the datum point (sometimes at an infinite distance from the region of interest) to the point in question. The earth's surface has also been used as the datum of potential. From the definition, it can be seen that the potential at the datum is zero. The unit of potential is the volt. The difference of potential between two points is the work done in moving a point charge from the first point to the second. It is equal to the difference in the values of the potentials at the respective points.

ELECTRIC POWER. Electric power is the product of electric current and electromotive force; that is, multiplication of current flowing by voltage forms the basis of the calculation of electric power. In a dc circuit, the current measured in amperes, multiplied by the voltage between wires, is the power in watts. A thousand watts constitute the kilowatt, a larger and more frequently employed unit of electric power.

The voltage and current may not be in phase with each other in an ac circuit and, while the instantaneous power is the product of the instantaneous voltage and current, this out-of-phase relation causes the power to fluctuate between positive and negative values. Hence for the average power (which is usually what is desired) this factor needs to be taken into account in determining electric power is an ac circuit, for it is only that component of the current which is in phase with the voltage that contributes to the average electrical power. The out-of-phase component produces the "wattless power." The power factor measures the fraction of the current that is in phase and available for true power. It is equal to the cosine of the phase difference between voltage and current. In a single-phase ac circuit having current of I amperes, voltage of E volts, and power factor f, the true power is EIf watts. In a balanced three-phase circuit, it is $\sqrt{3}\,EIf$ watts.

ELECTRIC POWER AND ENERGY MEASUREMENT. Over the years, the term *power*, in association with electricity, has tended to lose its true meaning. Thus, *power* is often found used in nontechnical literature where actually the correct term *energy* should be used. By definition, power is the rate at which energy is transformed or made available and is measured in *watts*. Energy may be defined as the time integral of power, or as the total energy supplied and is measured in *watt-hours*.

From an economic viewpoint, the most important of all electrical measurements is the measurement of energy. The watt-hour meter in various forms can be found in nearly every home, factory, highway billboard, and other locations where electrical energy is being purchased. Metering, installation and wiring have been governed by national, industrial, and local codes for so many years that at least in the United States, a particular type of installation is nearly identical everywhere.

For homes or small stores where energy demand is low and fractional horsepower motors are used, the common supply is single-phase, 3-wire. Two voltage levels are available, 120 and 240 volts, depending upon which pair of wires is selected. Electric ranges and heavy-duty home air conditioner motors can take advantage of the high voltage to reduce line currents, which reduces losses and permit smaller-size wiring.

Measurement of energy is almost always by means of fixed-installation metering. This provides safety through grounding of the meter enclosure and ease of reading through proper location and mounting. Tamperproof housings, which are also weatherproof where necessary, are common practice to insure the integrity of readings.

On the other hand, the measurement of power (watts) follows no such set of rigid rules. Very often considerable planning must go into a watt measurement to properly use existing metering or to purchase new equipment so that the test will be valid and results will be within the expected accuracy limits.

Whereas the measurement of energy is almost entirely restricted to 60 Hz, power measurements range from direct current to alternating current, including distorted waves, chopped waves, and missing pulses. A variety of circuits for connecting wattmeters to single-phase and poly-phase systems have been developed over the years. Basic connection diagrams appear in many electrical engineering textbooks. However, the user of a wattmeter is hard pressed to find diagrams covering practical or unusual situations. Wattmeter manufacturers usually offer an instruction book or a bound set of connection diagrams with their instruments. Basic wattmeter connection diagrams appear later in this article.

Power Theory. Since energy is simply the total power over a time period, an understanding of the power equations will provide some background into both power and energy terminology. A direct current circuit under steady-state conditions will produce power, computed as the production of the voltage across the circuit and the current in amperes in the circuit. This will also apply to alternating current circuits as long as instantaneous values of volts and amperes are used. The product of volts and amperes at any instant will give the instantaneous power in watts. However, such a measurement is unusual, difficult to make, and the resulting information is of limited use. Instantaneous power is of interest, of course, in the study of transient phenomena.

Average power in an ac circuit is of far more interest since it is equivalent to dc power and is a measure of mechanical work being done or heat liberated. Wattage or average power has an exact mathematical relation to horsepower or Btus or Calories.

The most basic equation for power, relating voltage, current, power, and the phase angle between the voltage and current, is derived as follows. If both voltage and current are sinusoidal, the average power over a cycle is:

$$P = \frac{1}{T}\int_0^T ei\,dt$$

$$P = \frac{1}{2\pi}\int_0^{2\pi} E_m \sin\theta \cdot I_m \sin(\theta - \phi)\,d\theta$$

where E_m and I_m are maximum values and ϕ is the phase angle by which the current lags behind the voltage.

From $\sin(\theta - \phi) = \sin\theta\cos\phi - \cos\theta\sin\phi$,

$$P = \frac{E_m I_m}{2}\left(\int_0^{2\pi} \sin^2\theta\cos\phi\,d\theta - \int_0^{2\pi} \sin\theta\cos\theta\sin\phi\,d\theta\right)$$

$$P = \frac{E_m I_m}{2\pi}\left[\left(\frac{\theta}{2} - \frac{\sin 2\theta}{4}\right)\Big|_0^{2\pi} \cdot \cos\phi - \left(\frac{1}{2}\sin^2\theta\right)\Big|_0^{2\pi} \sin\phi\right]$$

$$P = \frac{E_m I_m}{2}\cos\phi$$

The RMS values of sinusoidal voltage and current are:

$$E = \frac{E_m}{\sqrt{2}} \quad \text{and} \quad I = \frac{I_m}{\sqrt{2}}$$

Substitution in the previous equation yields:

$$P = EI\cos\phi$$

An immediate concern is what happens to the indications of a wattmeter if the voltage or current or both are not sinusoidal. Since it is possible to synthesize an odd wave shape with higher harmonics of the fundamental frequency, a wattmeter will give correct indications if it is frequency compensated over the span of harmonics. It is to be noted that if a particular harmonic is present in either the current or the voltage but not the other, it does not contribute to the average or active power. Frequency compensation of the wattmeter is still necessary for an accurate measurement.

Wattmeter Construction

Dynamometer. All wattmeters of this type contain a fixed coil (usually divided into two coils), which carries the current, and a moving coil having series resistance connected for voltage, turning within the fixed coil. The torque on the moving system is proportional to the product of the currents in the fixed and moving coils:

$$\text{Torque} = K_1 i_m i_f \frac{dM}{d\theta}$$

where M is the mutual inductance between the two sets of coils and remains constant over the usable scale angle.

The period of the instrument is very long compared with the period of the alternating voltage. Since the instrument movement cannot follow the rapid variations in torque, it will take up a balance position where the driving torque will equal the spring-restoring torque. Deflection represents the average torque:

$$\text{Deflection} = \frac{K_2}{T}\int_0^T ei\,dt$$

which is identical to the average power equation:

$$P = \frac{1}{T}\int_0^T ei\,dt$$

multiplied by a constant.

It follows then that a dynamometer wattmeter is a "true RMS wattmeter" and will take into account the magnitudes of voltage and current as well as the phase angle between them. Furthermore, meter indication will reverse when the flow of power reverses.

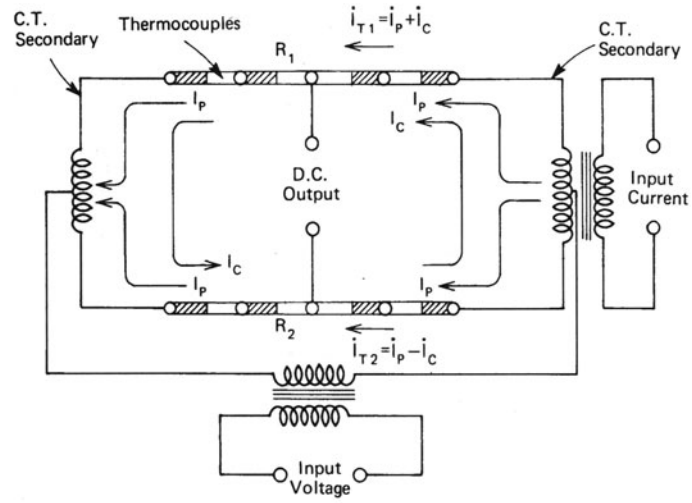
Thermal Watt Converter. As with the dynamometer, this type of instrument dates back to the early days of electricity. Heat produced by the voltage and current directly heats thermocouples which are arranged in a network to provide a dc output directly proportional to the wattage input. Figure 1 shows the essential parts of a thermal watt converter, namely, the potential transformer which is connected to monitor voltage, the current transformer which has a double-wound, center-tapped secondary, and the two sets of thermocouples.

The quantity of heat in $R1$ is proportional to $R1(I_{T1})^2$ and in $R2$ is equal to $R2(I_{T2})^2$. A vector diagram can be drawn for the sum and difference of I_p and I_c. From this diagram, equations can be developed which yield the difference in wattage in resistance $R1$ and $R2$ as $4RE1$ cos ϕ. The variable part of this term is EI cos ϕ which is the expression for ac power.

Some early thermal wattmeters used bimetallic elements or liquid-filled thermometers to measure the difference in heat between resistors $R1$ and $R2$. Thermocouples were used in more recent designs. In one design of a thermal wattmeter (Weston), the resistors and the thermocouples are one and the same. They act both as the heating resistors and as the temperature-sensing elements and show a very fast response to power changes. Also because the impedances of the several parts of the circuit are inherently balanced, there is little tendency for interchange of currents or potentials between the ac and the dc portions of the network.

Used mostly by the power industry, especially in totalizing of electric system loads, the watt converter has also found widespread use for measuring the power taken by very large motors where a remote readout is needed. Classed as a true RMS wattmeter, this device will respond to magnitude of voltage and current as well as the phase angle between them. Converter output will also reverse if the flow of power in the system reverses.

Specific Design Considerations. For clarity, it is believed best to comment pertaining to design factors pertaining to a few specific



i_p Current Due to Potential Transformer (P.T.)
i_c Current Due to Current Transformer (C.T.)
i_{T1} = Vectorial Sum of Currents in R_1
i_{T2} = Vectorial Difference of Currents in R_2
$R_1 = R_2$

Fig. 1. Basic circuit of single-element thermal watt converter.

commercial instruments, realizing of course that other designs are also available commercially. Available single-element and 2-element watt converters have a response to 99% in 0.7 second. Due to this rapid response, some protection of the thermocouples is necessary during overload conditions. Converters made by Weston use transformer cores which will saturate at moderate overload and minimize thermocouple burnout or other damage. If repeated overloads occur, the current circuit is usually operated below the 5-ampere normal level. Output is 50 millivolts open circuit per element when connected to a 500-watt load. An available 2-element watt converter has an output of 100 millivolts for 1,000 watts of circuit load. Provision is made for adjusting resistors in the output network so that the output can be reduced to achieve a particular ratio between input wattage and output millivolts.

Since a thermal watt converter has no moving parts to wear, it has been made very rugged by potting the entire circuit in a steel case. Polyphase wattmeters are easily made by assembling two or more single-element wattmeters and providing for a common output signal. An available 3-element wattmeter was designed for use on military motor-generator sets having 3-phase, 4-wire distribution systems. Such systems can be expected to operate under unbalanced conditions and to correctly read system power, a 3-element wattmeter must be used.

If the internal potential and current transformers are of good quality, the thermal watt converter can be used on frequencies extending to 20,000 Hz. A high-frequency type meter has a working frequency span of 180 to 20,000 Hz. In general, all thermal watt converters use internal transformers which excludes their use on direct currents.

The Low Power Factor Wattmeter

All of the so-called true RMS wattmeters will operate over the full range of power factor from zero to unity. Many low power factor examples can be found in the laboratory, such as motor or transformer testing, core loss tests, and power supply circuits. At zero power factor, a wattmeter will indicate zero watts even with rated voltage and current flowing. It quickly becomes apparent that the major difficulty in making a low power factor measurement is the low indication obtained on the meter. For example, if a wattmeter indicates full-scale with a unity power factor load, it will indicate half-scale with a 50% power factor load for the same level of voltage and current. At 20% power factor, the pointer will move only one-fifth the distance up-scale.

Special wattmeters are available for use on low power factor circuits. They are commonly called 20% power factor meters since the full-scale wattage is equal to the maximum voltage times the maximum current times 0.2. Both the accuracy and readability are improved through the use of this type of instrument. Since the 20% power factor wattmeter is designed to develop five times the torque of a unit power factor type instrument, care must be taken not to apply voltage or current above the maximum values shown on the instrument rating. Large overloads will soon burn out the resistors, fixed coils or moving coils. Small overloads continuously applied will cause deterioration of the overheated insulation. A wattmeter designed for low power factor can be used on higher power factor circuits provided either the voltage or current is sufficiently reduced to keep the pointer on the scale. Likewise, normal unity power factor meters may be used at low circuit power factors provided maximums are not exceeded.

Power Measurement

Power in an ac or dc circuit may be determined indirectly by making appropriate measurements of voltage, current and, where necessary, power factor. Power factor is usually expressed as a decimal value ranging between 0 and 1 and is derived from the cosine of the angle between the voltage and current. Power factor is further designated lead or lag, depending upon whether the current vector is ahead or behind the voltage vector based on counterclockwise rotation of the vectors. When making power calculations, it is not necessary to know if power factor is lead or lag.

Therefore, calculated power is a valid procedure, but with some reservations. Results are accurate only if all quantities of voltage, current, and power factor are correctly measured. The most elusive quantity is power factor. It is rare that a single-phase power factor meter can be used on other than 60 Hz. Even the 3-phase power factor meter has the requirement that the load must be balanced, although some designs have covered several thousand hertz. Modern electronic phase-angle voltmeters give excellent results on good sine waves over a wide frequency span. However, when distorted waves are encountered, results are questionable since many instruments of this type operate on the zero crossing principle. Further, the power factor of a distorted wave has little meaning since by definition it is based on a sine wave. The conclusion is soon reached that the only way to measure power accurately is through the use of a wattmeter.

In recent years, the phrase "true RMS wattmeter" has been reserved for the description of the ultimate in wattmeters. This is because some of the types of wattmeters available today will be accurate at only one frequency, must operate over a narrow voltage span, will not operate on dc, or must be worked at a high power factor. Although more descriptive than technically correct, the phrase will probably remain in the literature.

The original, basic, true RMS wattmeter was the dynamometer. Even until recently, this type of instrument was used as the standard wattmeter at the U.S. Bureau of Standards. A dynamometer wattmeter can be calibrated very accurately on dc and then used on ac. It is often the standard used to check other wattmeter devices because it can be made to a high accuracy, is a passive device, and will retain its accuracy for many years.

Probably the most important theorem in electric power measurement is that proposed by Blondel. In essence, it states that to correctly measure total system power, it is permissible to use one less wattmeter than current-carrying conductors. Also, the common point for the potential circuits is the conductor without a wattmeter current connection. The circuit being so measured may be operated at any power factor or condition of current or voltage unbalance. Strict adherence to Blondel's theorem would require the use of three wattmeters or a 3-element type meter to correctly measure a 3-phrase, 4-wire system. Since large commercial systems strive to maintain good voltage balance, a less expensive wattmeter of the $2\frac{1}{2}$-element design may be used and still achieve good accuracy.

Many questions often arise as to the proper wattmeter connections for various types of loads. For example, in a 3-phase, 3-wire circuit, the load can be delta, wye, or some other configuration. The wattmeter is only concerned with the three wires. This leads to a simple pictorial concept for the connection of a wattmeter. Visualize a laboratory bench with an unknown power supply on the left, the connecting wiring across the bench, and an unknown load to the right. Without knowledge of source or load, a true measurement of total system power can be made by following Blondel's theorem. Assume four wires are present and that all may be carrying current. Provide three wattmeters, making the wire without a meter the potential common. It is to be noted that one terminal of the meter current circuit and potential circuit carries an instantaneous polarity marking (usually plus or minus). That means that if (+) of a direct current supply is applied to each of these terminals, the meter will deflect upscale. Likewise if (−) were so applied, the meter will still go upscale. If one terminal is made the opposite polarity, the meter will move down-scale. In a multi-meter correction, the (±) current terminals should all be toward the source. Even so, due to load reactance, one wattmeter may produce a reversed indication. Total power then will be the algebraic sum of all meter readings. The measurement has been made without any knowledge of the source or the load. Voltage and current levels must be within the range of the meter to avoid overheating damage.

If more knowledge of the load can be obtained, the immediate benefit would be reduced metering costs. If we still have a 3-phase, 4-wire system, but know that the neutral wire carries no current, then only two wattmeters are needed to give a true reading.

If we further know that both voltage and currents are balanced around the phases, then metering costs can be further reduced by using a single wattmeter and a wye box. When combined with the meter, the two arms of resistance in the wye box form a wye having a neutral point. The wattmeter will then measure a phase power which is a known fraction of the total power. The scales of switchboard meters are usually direct reading, but the indication from a portable wattmeter must be multiplied by the wye box multiplying factor.

If only voltages in this 3-phase, 4-wire example are known to be balanced, then a so-called $2\frac{1}{2}$-element wattmeter is satisfactory.

Adherence to Blondel's theorem will always provide a true power measurement regardless of circuit conditions. A knowledge of the circuit will often point the way to less expensive metering which can provide adequate results.

For the measurement of power in a 2-wire circuit, Figures 2 and 3 show the possible connections. The connection shown in Fig. 2 indicating the wattmeter potential circuit connected on the load side of the current coils is most often used. Readings taken on small wattage loads may easily be corrected for meter loss by opening the load and reading

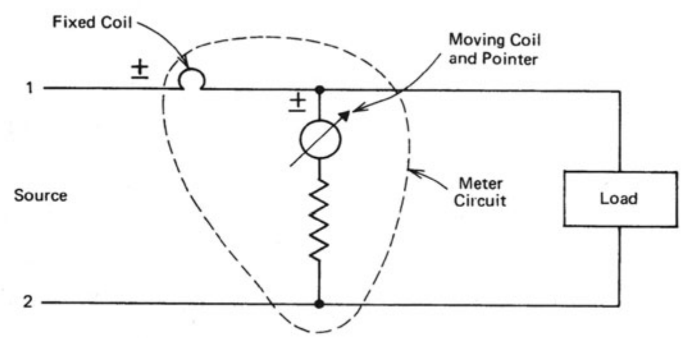


Fig. 2. Single-element wattmeter with potential circuit connected on load side. This connection is most often used.

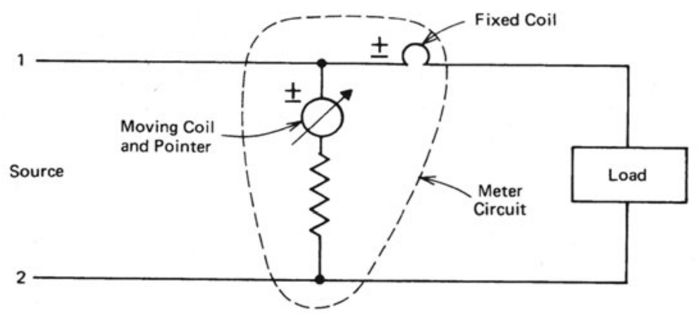


Fig. 3. Single-element wattmeter with potential circuit connected on source side. Usually used when fixed coil has a low current rating and a high voltage drop.

the wattmeter. Although not truly correct under varying loads, this "tare" reading will much improve the accuracy of the measurement.

If the wattmeter current coils are wound for low currents, such as 0.1 ampere, there is sufficient voltage drop across them under load to make the connection shown in Fig. 3 and thus yield a more accurate result.

Figure 4 clearly demonstrates Blondel's theorem of two wattmeters in a 3-wire circuit where the common potential circuit connection is made in the line without a current coil. There are some 2-element wattmeters available which connect the two moving coils into line 2. Such an arrangement is adequate for moderate voltages and accuracy, but the mechanical force set up between the fixed and moving coil due to the electrostatic effect precludes the use of this connection where high accuracy is needed.

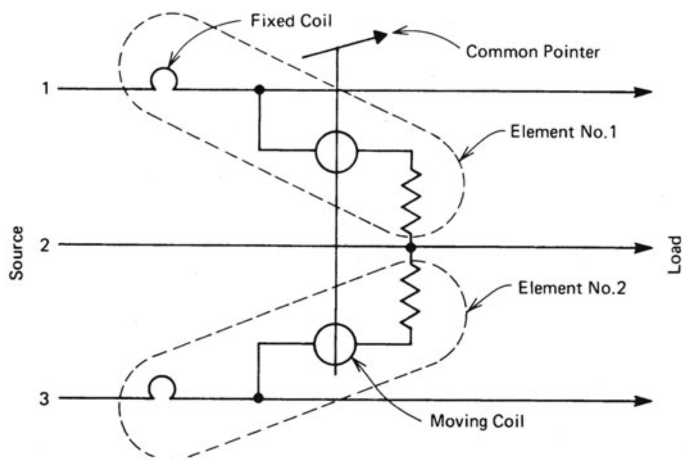


Fig. 4. Two-element wattmeter connected to a 3-phase, 3-wire system.

The $2\frac{1}{2}$-element wattmeter of Fig. 5 will monitor all the current flowing and so is capable of a correct wattage measurement where current unbalance exists. Line to line voltages should be nearly balanced and it is assumed that they are 120° from one another in vector rotation.

Power Factor

Whenever the voltage and current are not in phase, a third term called power factor must be introduced. The formula is expressed by:

$$\text{Power Factor} = \cos\left(\tan^{-1}\frac{Q}{P}\right)$$

where Q = reactive power, P = active power.

This basic formula may be used on simple, 2-wire circuits as well as a 3-phase system as long as the poly-phase system is balanced in both voltage and current.

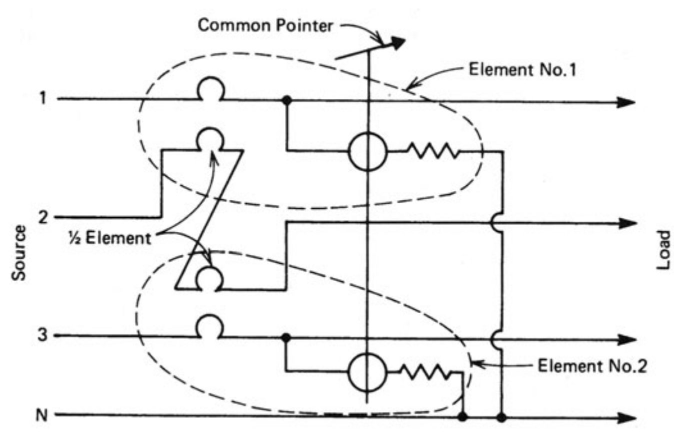


Fig. 5. A $2\frac{1}{2}$-element wattmeter connected to a 3-phase, 4-wire system.

When a poly-phase system is unbalanced in any manner, the system power factor will no longer have any specific physical meaning. A numerical ratio can be obtained and is defined as the interval power vector. This is not, in general, equal to the average value of the power factor during the interval.

Power factor is not usually one of the measured quantities when testing low power circuits under laboratory conditions. Commercial users of bulk power very often monitor power factor so as to be able to make adjustments of condenser banks or synchronous motors to keep the power factor as high as possible. This, in turn, usually reduces the cost of the power purchased.

Power Factor Meter. Since the power factor meter can show at a glance the operating condition of a power system, it is most often found on the switchboards of both consumer and supplier of power.

Ratio type movements are commonly found in both single phase and poly-phase power factor meters. The single-phase power factor meter measures the ratio of vars to watts which corresponds to the tangent of the power factor angle when the voltage and current are sinusoidal. A scale can then be drawn for the power factor which is the cosine of the angle.

The poly-phase power factor meter also uses the same basic ratio mechanism as the single-phase meter and is connected 3-phase, 3-wire. This instrument will indicate vector power factor by measuring the angle between line current 2 and line voltage 3–2 and 1–2. The indication is the poly-phase power factor only for balanced voltages and currents when both are sinusoidal. Some years ago, Weston offered a poly-phase power factor meter for use on unbalanced systems. However, due to limited acceptance because of the high cost resulting from the complexity of construction, it was discontinued. It is doubtful that such a meter can be purchased today from any manufacturer.

Single-phase instruments can be scaled in a variety of combinations, such as 0–1 P.F. lag or lead, or 0.3–1–0.3 P.F. Due to the principle of the instrument, not every range combination is possible in the poly-phase power factor meter.

When compared to the other electrical quantities, the *var* is a relatively new term, having been recognized by international agreement in 1930. The letters were taken from volt-ampere-reactive and represent power incapable of producing work. Voltages used in this form of metering are always 90° in vector rotation from that used in wattage measurement. In any alternating current system having sinusoidal voltages and currents operating at other than unity power factor, the real power is less than the volt-ampere-product and is related by the familiar right triangle. See Fig. 6.

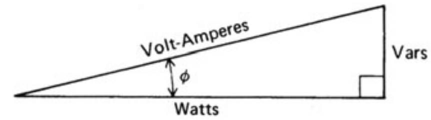


Fig. 6. Real power is less than volt times amperes. ϕ = power factor angle.

Since more iron and copper are required to deliver a given amount of power at a low than at a high power factor, design allowances must be made for any reactive volt-amperes in addition to the designed load power. Line losses are higher and voltage regulation poorer when the power factor is low. Any customer contributing to a low power factor must be expected to pay for this added loss in addition to the actual energy consumed. Var metering is generally used to obtain an estimate of the average power factor of a fluctuating load over a period of time. A recording type meter would be used for this measurement.

$$\text{Var} = EI\sin\phi = EI\cos(90° - \phi)$$

A wattmeter can be converted to a varmeter if the voltage is shifted into quadrature with the line voltage at the load. In single-phase varmeters, this voltage shift is done by means of added reactance. In poly-phase systems, the voltage shift is most easily done by reconnection if all necessary points are available. On 3-phase, 3-wire systems, a var connection may be made to wye-connected resistors forming an artificial neutral. A special poly-phase, phase-shifting autotransformer, also called a potential converter or phasing transformer, is available for var service. When a varmeter is purchased from the manufacturer, the proper scaling, instrument adjustment (watts), and connection diagrams are provided. If it is desired to convert an existing wattmeter to a varmeter, several problems must be avoided (or at least understood).

Varmeter indication can be either up- or down-scale, depending upon whether clockwise or counterclockwise rotation is selected for the new var voltages. Either is correct, but it should be determined that up-scale indication will occur for a leading or lagging power factor. Another more serious problem is that of voltage level. If the new var voltages differ from those used by the basic wattmeter, then, for the same power supplied to the load, the full-scale value and therefore all scale points will be in error. In order to correct this, the meter series resistance would have to be changed (a task for a meter shop with wattmeter calibration equipment).

Varmeters are usually arranged to deflect to the right (up-scale) on leading power factor circuits with phase rotation 1-2-3. Zero center varmeters also have the words IN at the left and OUT on the right. The OUT refers to the flow of power from a source, considered to be located at the left of a diagram, to a load somewhere to the diagram right.

Instrument Transformers

For reasons of personnel safety and good instrument design, it becomes impractical to connect meters directly to circuits having voltages above 1,000 volts and currents above 50 amperes. Furthermore, instruments tend to become inaccurate when directly connected to high voltage because of electrostatic forces that act on the moving system. Shunts are often considered for large ac currents, but have the disadvantages of large power loss at high current ratings and no voltage isolation. Unless a shunt is especially designed for ac use, its inductance, nonuniform split of current in the blocks, skin effect, proximity effect, and nonuniformity of the material all contribute to ac shunt error.

By contrast, potential and current transformers offer a practical and safe means of reducing voltage and current by an exact ratio for instrument use. All instrument transformers have the following basic design objectives: (1) careful attention to an exact primary-to-secondary ratio; and (2) as small a phase angle as possible. Instrument transformers by design have small load (burden) capability because meters do not need large amounts of power and a low burden will enhance the design for best possible accuracy.

In a power station, the instrument transformers which are used for station metering, relay operation, and control services are often several feet tall (a few meters), resulting from a design to withstand very high voltages. For laboratory and shop testing at low distribution voltages, both potential transformers and current transformers are very small in comparison. They weigh perhaps 15 pounds ($\sim$7 kilograms) and can be hand carried. Quality potential transformers for laboratory use will support a 25-volt-ampere burden, have a ratio accuracy from 0.1 to 0.5%, and a phase angle of 10 minutes. A core-type design in which the winding surrounds the iron provides the necessary insulation for a high range of 2,300/1,150 or 115 volts. The iron is usually grain oriented silicon steel from 0.012 to 0.025 inch (0.3 to 0.6 millimeter) thick, depending upon the quality of the transformer and frequency range. Current transformers of the toroidal type, using tape-wound, high nickel-iron cores, have phase angles of less than 2 minutes and a ratio error of plus or minus 0.02%. Burden capability is up to 25 VA. Primary-to-secondary insulation is difficult to provide in a toroidal transformer. A rating of 2,500 volts is common for stock transformers with 5,000 working volts pushing the practical limit for this construction in custom designs.

Automated, Remote Meter Reading. The concept of using meter-mounted sensors connected to an information network, such as the telephone system, was initially suggested in the 1960s. However, little progress has been made until recently, and this is in connection with selected industrial/commercial users of a large amount of power. As of 1993, this program has passed the field testing stage. In the mid-1980s, a pipeline company complained to a large southeastern U.S. electric utility that the utility's billing cycles, which had different start and end dates in different locations, were not synchronized with the user's monthly pumping cycles. As a consequence, the pipeline firm had difficulty in managing the cost of electricity required to move petroleum products through its vast pipeline network. With energy costs representing 40% of the company's annual operating expenses, this was no small concern. The solution was *advanced metering*.

Through this new program, known as ROCS™ (Read-Only Central Station), large commercial and industrial customers can access, by computer, real-time information from their electric meters. With such data, users can take advantage of cheaper rates while monitoring and managing energy consumption. In the system, each customer provides its own personal computer and a telephone line to connect the PC to the meter. A video display monitor, a modem, and proprietary software also are required. Users may examine their data privately at any time, without involving the utility. Security is maintained through multiple password protection, hardware security keys, and the limited read-only functions of the software, which ensures the operating integrity of the recorder. Users can monitor and audit daily energy use at multiple locations and predict their energy billings.

Additional information can be found in the Timberman reference list at the end of the article on **Electric Power Generation and Distribution.**

ELECTRIC POWER PRODUCTION AND DISTRIBUTION.

Few people would disagree that an abundant, reliable, and comparatively low cost electric power supply is fundamental to the maintenance and growth of the world economy, especially in the case of the advanced, industrialized nations.

Over the years, since the beginnings of centrally produced and distributed electric power (1880s), the major problem facing electric utility managers, whether investor or publicly owned, was simply that of *keeping up with the demand.* Utility managers were among the first industrialists to establish long-term planning, required because of the long lead times imposed between the design phase and putting a new plant or expanded facility onstream. Lead times became even longer in the case of nuclear power plants, ranging from 5 to 8 and even 10 years from blueprints to delivery of power.

Today, electric power planning has become even more complex as the industry grapples with the problems of environmental protection. The introduction of effluents into the air as the consequence of combusting fossil fuels to create steam and hot gases to operate the electricity-generating turbines poses severe technical complexities, either in terms of (1) treating and neutralizing the undesirable effluents (principally sulfur dioxide (SO_2), nitrogen oxides (NO_x), and carbon dioxide (CO_2), as well as solid particulates), or (2) essentially treating or transforming the raw fuels used. Coal, in particular, is implicated in connection with SO_2, but nitrogen oxides and CO_2 are natural products of combustion, thus implicating all fossil fuels. By comparison, it is interesting to note that these problems are not present when nuclear fuels are used.

Because of the technologies needed to cope with environmental problems, electric power planning today is additionally faced with selecting one solution for a given problem out of a choice of several proposals. (Although the results desired from technology can be legislated, the actual achievement of results is not precisely programmable.) The complex politico-scientific formulas currently applying to coal-fired power plants in the United States are described in the article on **Coal Conversion (Clean Coal) Processes.**

With an expected doubling of world population to occur sometime in the next century, electric power planners are also greatly concerned with how the supply can keep up with what appears to be phenomenal demands if current standards of living are to be maintained in the present industrial nations and improved in the developing nations. This latter topic is beyond the scope of this article, but reference is made to an excellent scholarly article by Robert Fri (listed at end of this article).

Emphasis on Research

If any nation is to assure a reliable electric power supply now, within the next few decades, and beyond, the answers must derive from advanced technologies, some of which are on hand and are being implemented, some of which still are essentially in their infancy, and some of which are as yet unknown. Thus, electric power planning now must include inputs from costly research and development programs that not only tackle the near-term problems, but also peer far into the future. This is a task that is national (and international) in scope and far beyond the capability of any one electric utility. Thus, it was a laudable exercise of logic when a few years ago over a hundred electric utilities joined together to form the Electric Power Research Institute to concentrate on finding and developing technical solutions to present and future problems. Similar but smaller efforts of this nature also have been undertaken by some of the other very advanced industrial countries. These concerted research efforts are described later in this article.

Electric Power Statistics

Global statistics of electric power production are difficult to compile and thus are not fully reliable. For example, statistics from the former Iron Curtain countries were always subject to question. For general purposes, however, the percentages shown in Fig. 1 should suffice.

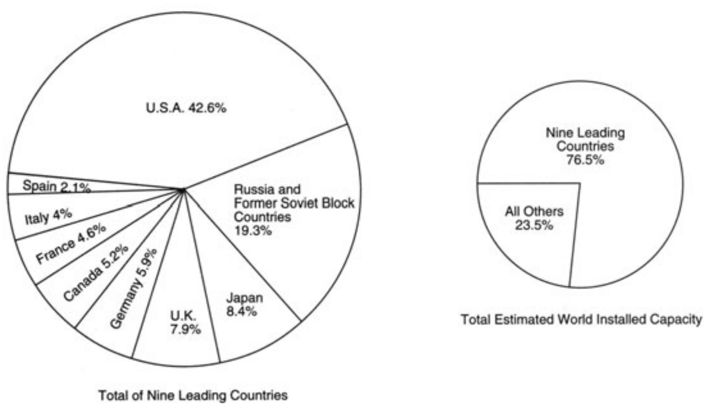


Fig. 1. Distribution of electric generating capacity among leading countries.

According to a forecast (1991) made by NERC (North American Electric Reliability Council), the electrical peak demand in the United States will grow at a rate of 1.9% per year from 551,500 MW (megawatts) in 1991 to 651,100 MW in 2000. In the United States, peak demands occur in summer. See Fig. 2. The United States is anticipated to continue to be summer peaking through 2000. The summer peak demand is about 11% greater than the winter peak.

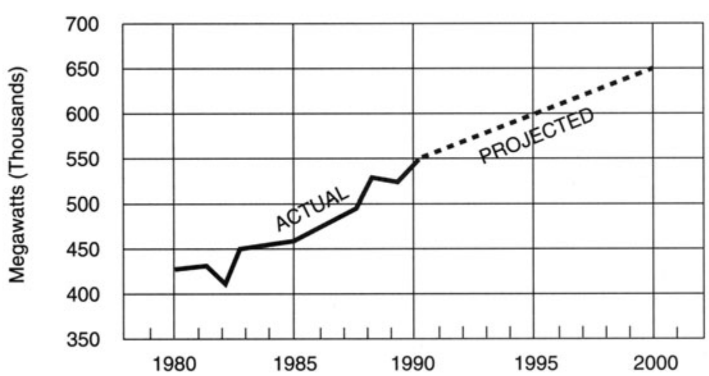


Fig. 2. Actual and forecasted summer peak electric power demands in the United States. (*Adapted from North American Electric Reliability Council.*)

The peak demand in Canada, which occurs during winter, was 80,900 MW in 1991 and is projected to be 97,100 MW in 2000, representing a growth rate of about 2% per year. Canada's winter peak is projected to exceed its summer peak by about 40%. See Fig. 3. The peak demand and annual net energy for load projections for the United States and Canada are given in Table 1.

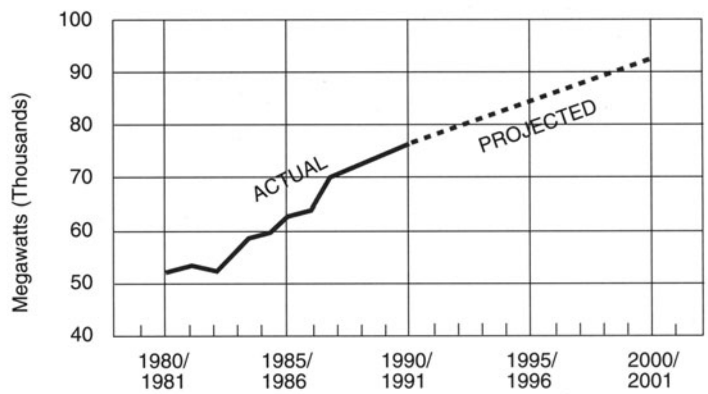


Fig. 3. Actual and forecasted winter peak electric power demands in Canada. (*Adapted from North American Electric Reliability Council.*)

TABLE 1. PEAK DEMAND AND ENERGY FORECAST
Electric Power in United States and Canada

	1991	2000	Growth (%/yr)
Peak demand (thousands of MW)			
United States (summer)	552	651	1.9
Canada (winter)	81	97	2.0
Net energy for load (millions of MWh)			
United States (annual)	2951	3513	2.0
Canada (annual)	446	548	2.3

A typical peaking curve for a U.S. utility (summer peaking) is shown in Fig. 4.

The electrical power requirements or load placed upon a specific power-generating facility may, at times, be greater than or less than the optimal rate in terms of economy, profit, and other operating factors. Within the upper and lower limits of capacity of any facility, operating adjustments are made, cutting in generating equipment or cutting out units, as may be required. But it may be more economical for a facility to buy power from another generating facility or, depending upon the direction of the demand, to sell power to another facility. This can be accomplished smoothly by means of a tie-line, or, if large numbers of facilities are so connected, the system may be part of an *interconnecting*

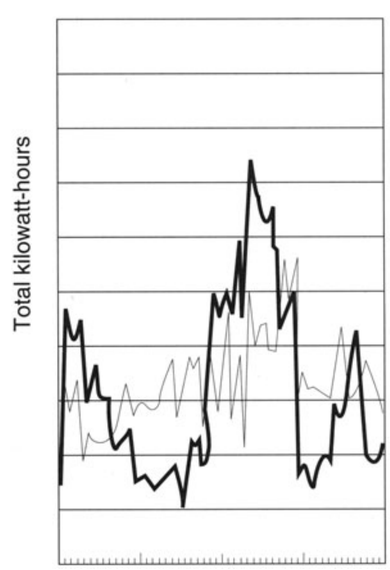


Fig. 4. Annual peaking curve for a typical U.S. electric utility. Heavy line is total as reported. Fine line is seasonally adjusted.

network. In the latter situation, not only are individual facilities connected, but whole utility systems may form a vast network encompassing very large geographical areas. In some countries, all or nearly all facilities are so interconnected. In the case of North America, there are interconnections between two countries—the United States and Canada.

Fundamentally, networks are established in the interest of maximizing operating economy and for providing extra power when required, as may be caused by any number of factors, such as storms, which may affect one or two generating facilities in a given area quite seriously. In fact, the public is made aware of such interconnections in times of emergencies. Power from other sources is conveniently available, too, to handle scheduled maintenance shutdowns.

T. J. Nagel has described the situation succinctly: "The purpose of interconnections is to expand the scope of the individual power systems so as to enhance both reliability and economy of the power supply. In terms of reliability, interconnections provide assistance during generation outages, assist in distributing excessive generation at times of major load outages, and provide support in times of transmission outages. In economic terms, interconnections allow the full exploitation of economies of scale in both generation and transmission. Savings thus can be realized without compromising reliability, through the sharing of risk. Interconnections allow the interchange of power to reduce generation costs". With reference to a 1991 report prepared by NERC, it is observed that, in the 100-year history of the electric utility industry in North America, utilities have grown from isolated systems supplying their customers from local generating sources to a highly interconnected system of generating stations and high-voltage transmission lines encompassing large geographic areas and many individual utilities. As shown by Fig. 5, in NERC, there are four major interconnection areas: (1) the Eastern Interconnection, (2) the Western Interconnection, (3) the Texas Interconnection, and (4) the Québec Interconnection.

Within each interconnection, individual utility systems are operated in synchronism with each other. Within the limits of the transmission systems, demand at any point in an interconnection can be supplied by generation located at any other point. Transfer of electricity from one point to another will, to some extent, flow over all transmission lines in the interconnection, not just those in the direct path of the transfer.

Fig. 5. Four major interconnection regions. (*North American Electric Reliability Council.*)

The ability to transfer electricity *between the interconnections* is very limited and depends on the capacity of the *tie-lines,* the ability to deliver electricity through the network to the tie-points, and the amount of generating capacity available at the time of the transfer.

Generating Capacity within NERC. New unit capacity additions of 113,400 MW are planned in NERC over the 1991–2000 period. A summary of these generator unit additions, changes, and facility retirements is given in Table 2.

Some major interconnections are comprised of two or more subregional bodies, as indicated by the map in Fig. 6. Recent forecasts indicate that the period through 2000 may experience relatively small mar-

TABLE 2. GENERATING CAPACITY PROJECTIONS
Aggregate 1991–2000 Forecast—United States and Canada
(Additions, Changes, Facility Retirements)

	United States		Canada		NERC-Total[1]	
Type	Thousands of MW	% of Total	Thousands of MW	% of Total	Thousands of MW	% of Total
Nuclear	5.9	6.6	2.6	11.5	8.6	7.5
Coal	15.1	17.0	1.9	8.3	17.7	15.6
Hydro	1.1	1.2	12.7	55.0	13.9	12.3
Oil/gas	34.2	38.3	2.3	9.9	36.8	32.5
Pumped storage	2.5	2.8	0	0	2.5	2.2
Other (utility)	12.8	14.3	1.1	4.8	13.9	12.2
NUGs (net additions)	17.7	19.8	2.4	10.5	20.1	17.7
Additions	**89.3**	**100**	**23.0**	**100**	**113.4**	**100**
Changes	10.6		0.4		11.0	
Retirements	−4.3		−0.5		−4.8	
Net Total	**95.7**		**22.9**		119.6	

[1]Includes additions in WSCC-Mexico.

Utility oil-fired (4300 MW) and gas-fired (29,900 MW) units account for 38.3% of the total 89,300 MW of planned additions in the United States. The second largest additions in the United States are the nonutility generators, at 17,700 MW, or 19.8%, nearly half of which are gas-fired. A significant number of nonutility generators (NUGs) exist in California, Texas, and New York. The foregoing projections show major increases from NUGs in NPCC, SERC, WSCC, and MAAC.

In Canada, hydro units account for 55%, or 12,700 MW, of the 23,000 MW of total planned additions over the forecast period.

(SOURCE: *North American Electric Reliability Council.*)

Fig. 6. Regional reliability councils that make up the total NERC network. Legend: ECAR (East Central Area Reliability Coordination Agreement); ERCOT (Electric Reliability Council of Texas); MAAC (Mid-Atlantic Area Council); MAIN (Mid-America Interpool Network); MAPP (Mid-continent Area Power Pool); NPCC (Northeast Power Coordinating Council); SERC (Southeastern Electric Reliability Council); SPP (Southwest Power Pool); and WSCC (Western Systems Coordinating Council. Also, not shown, is an affiliate, ASCC (Alaska Systems Coordinating Council).

gins for contingencies, or are likely to need capacity resources in addition to those currently planned. Regional-specific contingencies and concerns include risks of higher-than-expected peak demand growth, lower generation availability, or the inability to improve it, especially during peak periods, or the reduced effectiveness of projected demand-side management programs. Areas with the smallest amounts of capacity available for contingencies are in the Eastern Interconnection. This interconnection also is the most susceptible to potentially adverse effects from the U.S. Clean Air Act Ammendments of 1990. In addition, a series of Clean Air Programs has set tight goals to drastically cut emissions.

Electric Power Network Research. From a fundamental standpoint, the great bulk of instantaneous support to make up a generating deficiency in some part of a network is derived from the inertial energy stored in the combined rotating masses of all generating plants and other electricity resources in the network. Although such a transfer may be accompanied by a slight, very temporary decline in frequency in the network, this is not a critical factor in making the transfer. The much greater role in sustaining continued service and reliability is the capability of the transmission facilities to handle the resulting shift in power flows. Basically, the flows of power on individual elements of an interconnected network are usually uncontrolled. They are distributed in accordance with the impedances of the circuit elements. Consequently, if these elements are overloaded and disconnected by protective equipment, uncontrolled power failures in some parts of the network may occur. Over the years, with particular emphasis since the massive electric power failure that blacked out most of the northeastern United States and parts of two Canadian provinces on the night of November 9–10, 1965, many studies, including sophisticated computer-assisted programs and simulations of emergency situations, have been made. The blackout of 1965 and a subsequent blackout of New York and environs in 1977 were thoroughly investigated and documented. The chain of happenings, including equipment malfunction and network management communications, are beyond the scope of this article.

The transmission system, of course, behaves according to the basic laws of electricity. The amount of power flowing over any one line is determined by Ohm's law (current flow is directly proportional to voltage and inversely proportional to impedance). The principles involved are shown in Fig. 7. The Kirchoff laws of network operation (the voltage around a closed loop must sum to zero, and currents into a node must sum to zero) also apply.

Electric Power Consumption

Some detailed studies of electric power consumption have been made by regional utilities, but the suppliers of power are more concerned with the impact of new users. With so much geographical decentralization of manufacturing firms occurring today, the impact of new industries must be determined well in advance. The information given in Fig. 8 represents an average usage in the United States over a number of years. Generally, this mix will apply to other advanced industrialized countries. Also see Fig. 9.

Raw Energy for Electricity Generation

Numerous raw energy sources have been considered over the past century when conveniently available electric power became commonplace. Wood and coal were used initially, and coal continues to be in demand because of the vast reserves and the number of existing plants that are designed for coal-firing. Some of the environmental problems in connection with coal and other fossil fuels were mentioned earlier in this article. Also, there is a separate article on each of these major fuels elsewhere in this encyclopedia. For example, see **Coal; Combustion; Fuel Cells; Geothermal Energy; Natural Gas; Nuclear Power; Petroleum; Solar Energy;** and **Wind Power.**

Intermediate-term Energy Supply Outlook (1990s–2000). The North American Reliability Council observed in 1991 that new issues are emerging regarding the long-term adequacy of the supply and infrastructures associated with coal, natural gas, and fuel oil (petroleum).

Coal. Adequate supplies of low-sulfur coal may not be available in a timely manner to meet the increased demands caused by Clean Air legislation. The increased demand is likely to strain low-sulfur mining capacity in the eastern United States as well as the transportation systems in both the East and West. The mining capacity for low-sulfur eastern coal may not be adequate, and this could lead to significant use of Western coal in the eastern and midwestern NARC regions. Use of low-sulfur western coals in place of eastern coal will pose significant problems of transportation and supply.

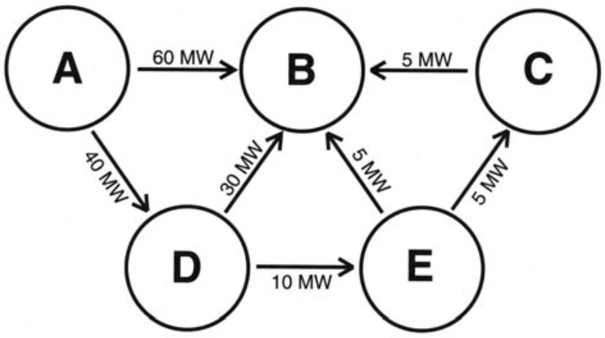
Several ways to meet the SO_2 emission limitations of the Clean Air Act are available and include:

Changing the fuel in affected units;
Switching to lower-sulfur coal, lower-sulfur oil, or natural gas;
Removing SO_2 after combustion by scrubbing, or possibly removing sulfur using clean coal technologies; and
Retiring facilities where compliance costs exceed a unit's value.

Natural Gas. Limitations currently exist in the natural gas delivery system that constrain the use of gas for electricity generation, particularly during winter. These limitations are likely to become more critical. Some of the developing regions of NARC are remote from natural gas supplies. Efforts are underway to improve natural gas delivery. For example, the import of nearly 400 million cubic feet per day of Canadian natural gas into the northeastern United States has been approved. The gas will be carried by a new 370-mile (596-kilometer) pipeline from Canada into Connecticut and New York. Many electric utilities routinely switch from natural gas to fuel oil in November, as increasing heating demands for natural gas strain the delivery system. Additional concern is expressed where a single pipeline is involved. Although pipelines are highly reliable, a pipeline failure could have a heavy impact on the electric utility industry in the area affected. See Fig. 10.

Petroleum. Two major types of oil are used for electricity generation: No. 6 residual fuel oil, and No. 2 distillate fuel oil. Environmental constraints will require the use of low-sulfur fuel oil. The U.S. refining capacity for low-sulfur residual oil is limited because it is less profitable and the availability is not likely to increase. In some states, regulatory agencies have placed limits on the amount of fuel inventories allowed in utility rate base, usually on the basis of historical consumption. However, inventories are meant to meet needs during supply disruptions that could be more severe than historical occurrences. This is particularly true when oil is used to fuel peaking capacity or as a backup fuel for natural gas. See Fig. 11.

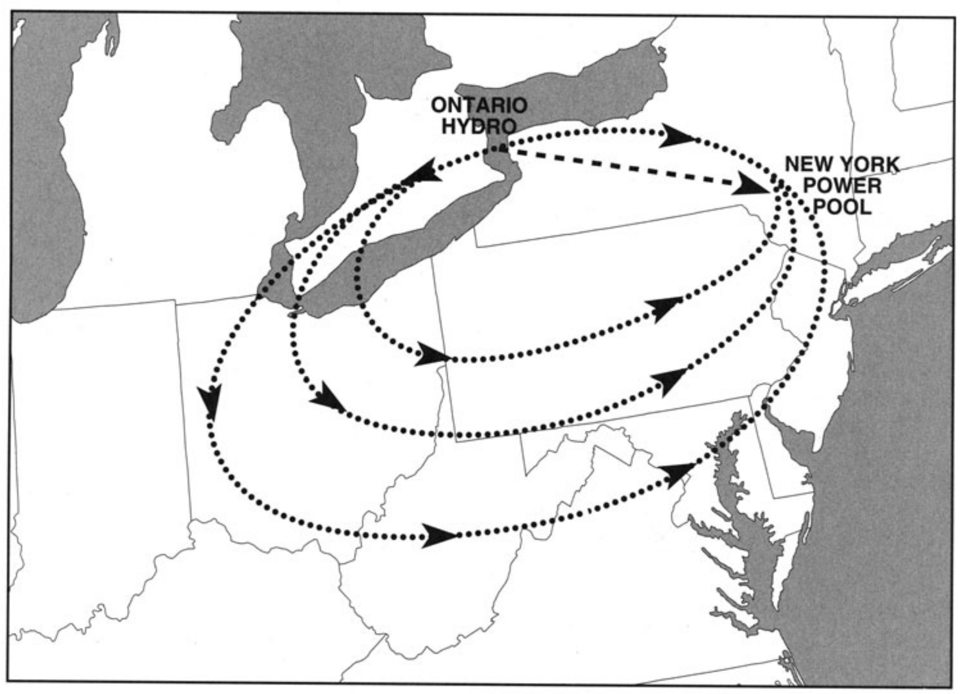
CONTRACT → B 100 MW
ACTUAL



Total Current (A)	Increment of Additional Current (A)	Total Losses (A)	Incremental Losses (A)
100	–	10,000	–
200	100	40,000	30,000
300	100	90,000	50,000
400	100	160,000	70,000
500	100	250,000	90,000

In table, A = amperes.

(a)



(b)

Fig. 7. A large electric power exchange (interconnection) network follows the laws of Ohm and Kirchoff. (a) Schematic diagram of an interconnected power system. If Utility A wants to sell 100 MW to B, power may flow to accomplish this transaction as shown. The 100 MW will not flow directly from A to B, but it will divide according to the impedance of the system. The current will not take the single path of least resistance, but it will divide in a way that is inversely proportional to the impedance of all the paths from A to B, thus obeying Kirchoff's network laws. Now, if D wants to sell to E, the two parties must take into account the flows resulting from the transaction between A and B. If the circuit from D to E is loaded to its maximum by the flow introduced from A to B, then D and E cannot complete their transaction. This loop flow limits the use of the transmission system.

(b) Large geographical scale network, illustrating a purchase of power by the New York Power Pool of 1000 MW from Ontario Hydro in Canada. Only half of the desired purchase flows directly to New York. Lower network impedance in Michigan, Ohio, and other states causes a significant amount of the power to flow through these states. Loop flow is well known to utility system operators. It registers on their meters and is calculated in advance, but it does limit the ability of systems through which it flows to make additional transfers. (*Electric Power Research Institute.*)

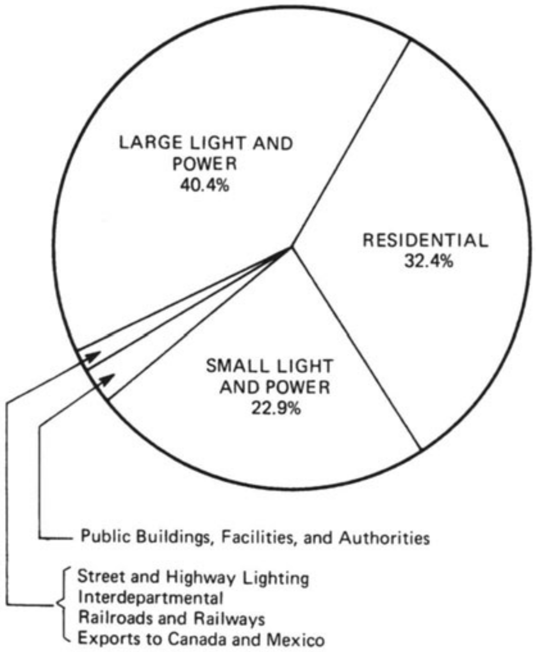


Fig. 8. Distribution of electric power requirements in the United States (average of several years).

Electric Power Generating Systems

The use of rotating equipment for generating electric power dates back well over a century. See Fig. 12.

The majority of electric power produced in the United States is by 3-phase generators. Advantages of 3-phase generators lie in economy of apparatus, lower transmission losses, inherent starting torque for polyphase motors, and a constant running torque for balanced loading. A generator is built with axial slots for armature coils in a stationary hollow cylindrical iron core called the stator. The windings are placed in the slots so that when carrying current they produce a chosen even number of alternate magnetic poles. The coils over each magnetic pole are grouped in three equal bands to give a 3-phase balanced system of terminal voltages.

An inner rotor has coils which carry direct current to give the same number of alternate magnetic poles as on the stator. Rotor current strength is controlled by a rheostat or voltage from a direct-current generator. Voltages are produced in the stator windings by flux cutting as the rotor magnetic flux sweeps by them, and currents flow when the generator terminals are connected to a 3-phase load impedance. The 3-phase stator line voltages are equal in magnitude and 120 electrical degrees apart in time sequence. So also are the line currents for a balanced 3-phase load. Generator voltages are of the order of 12,000 to 30,000 volts for large machines.

Generator frequency is the product of the pairs of magnetic poles and the speed in revolutions per second. At 60 cycles (60 Hz), a 2-pole generator runs at 3,600 revolutions per minute; a 6-pole generator at

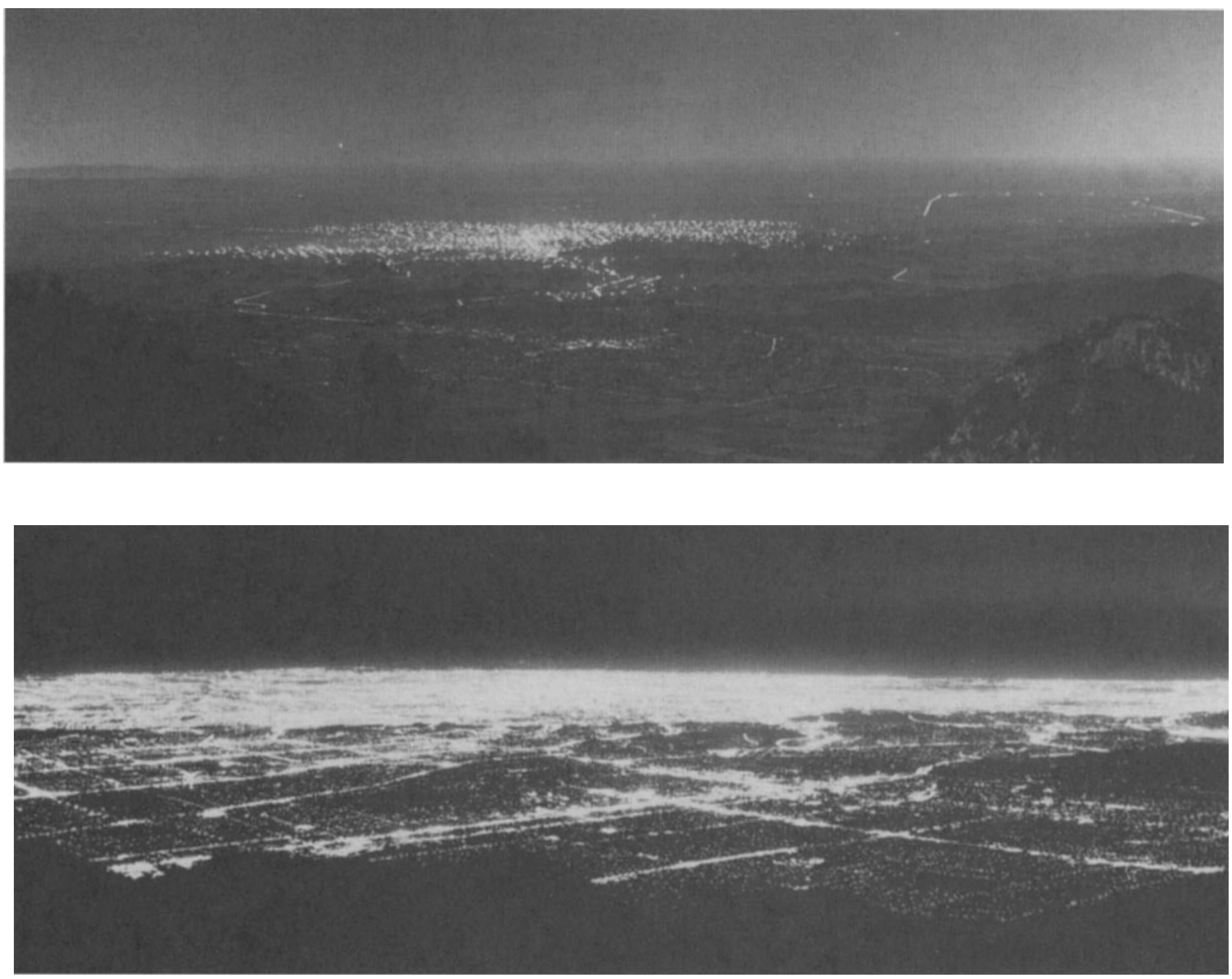


Fig. 9. The fantastic growth of electric power that has occurred during the present century is dramatically illustrated by these two views. They illustrate, of course, only one of the numerous major uses of electric power, illumination. (a) View of the Los Angeles basin from Mount Wilson Observatory taken in 1908; (b) the same view taken in 1971. (The effects of massive city lighting on telescope viewing are discussed in the article on **Light Pollution.**) (*Kitt Peak National Observatory.*)

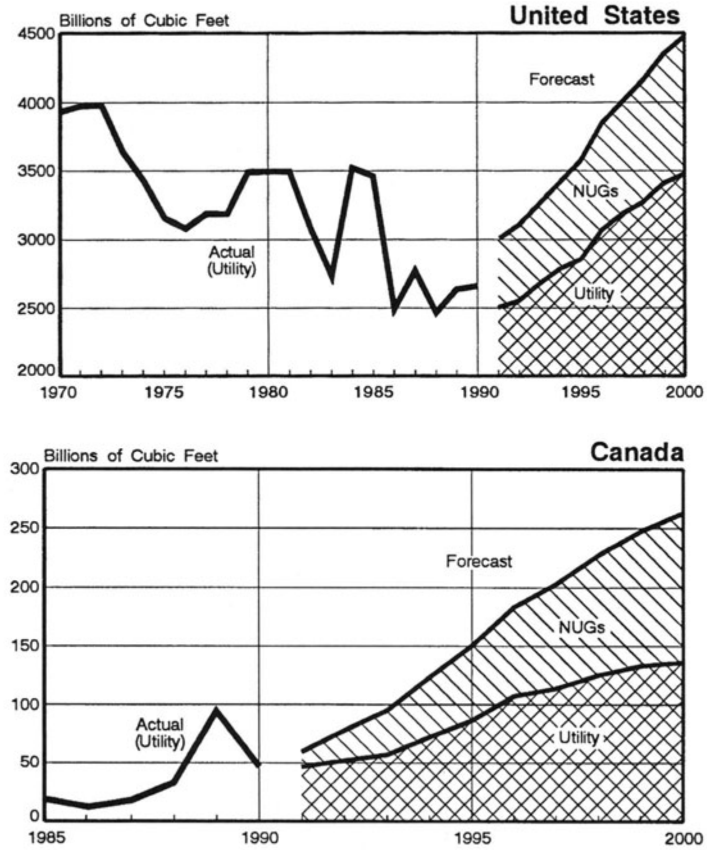


Fig. 10. Actual and forecasted consumption of natural gas for generating electricity. NUGs = nonutility generators of electricity. (*North American Electricity Reliability Council.*)

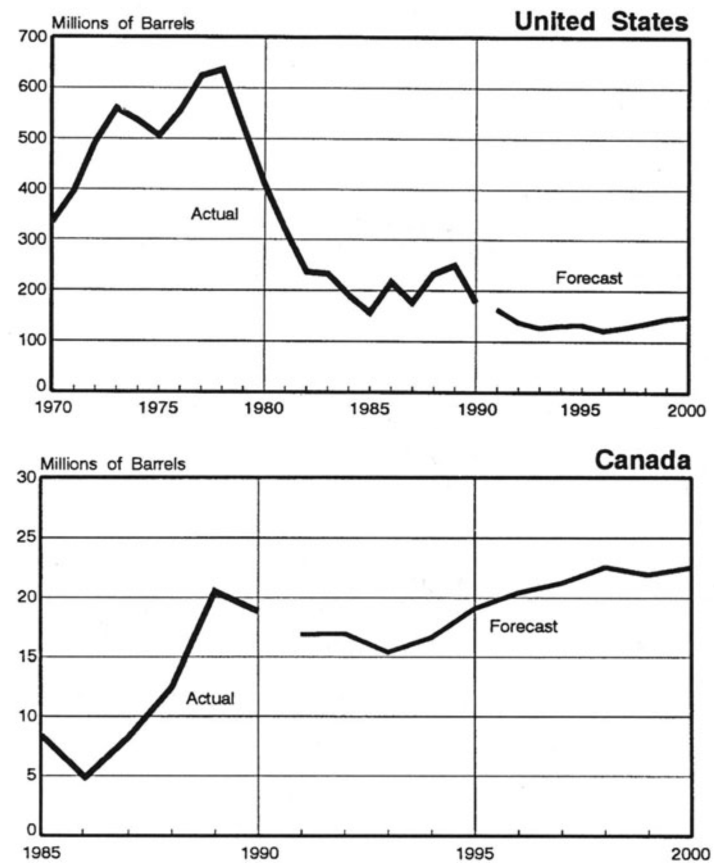


Fig. 11. Actual and forecasted consumption of fuel oil for generating electricity. Nonutility generators of electricity are not included in these charts. (*North Americn Electricity Reliability Council.*)

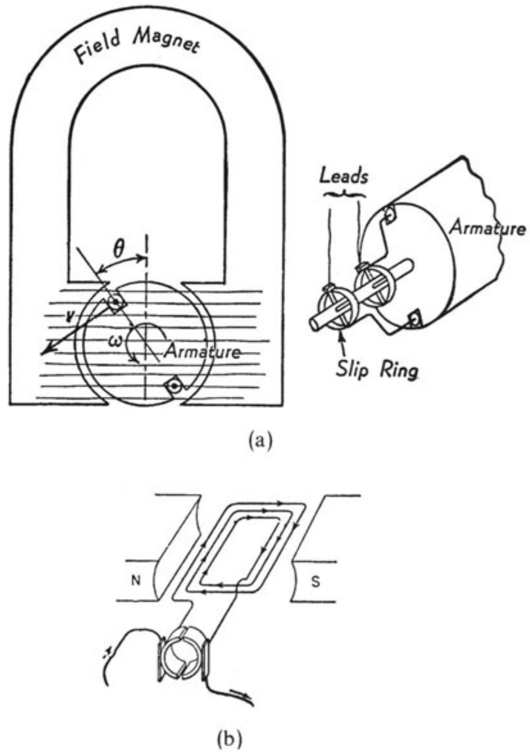


Fig. 12. Simplistic view of electric generator that consists of a soft iron core rotating between the poles of a permanent magnet, and having the slots on the surface, in which is embedded a coil of wire. It is apparent that as the coil rotates, carried by the soft-iron armature, it will cut across the flux lines, extending from pole to pole as shown in (a). When this apparatus is connected to stationary leads through the medium of slip rings, and brushes resting thereon, it becomes an elementary alternator. If, instead of slip rings, a split segment, such as that shown in (b), is connected to the ends of the coil, the reverse of the current in the coil will occur when alternate segments of the slip ring (an elementary commutator) are opposite one of the brushes. This gives unidirectional current in the exterior leads, although it would be quite variable with only one coil in the armature.

A uniform and unidirectional current is the result of many single-coil armatures so connected that the resultant current is the sum of several individual outputs. With sufficient overlap, the resulting current will be uniform and unidirectional. When the coil is revolving at a speed of ω radians per minute, at the position indicated, the speed of cutting vertically across flux lines is $v \cos \omega t$. The time is measured from the vertical position of the coil, and angle ϕ is ωt. When a wire cuts a magnetic field having a flux density represented by b and has a length and velocity represented by l and v the voltage generated is $blv/10^8$ volts, thus the voltage generated any instant t, t being measured from the point of minimum generated voltage, is $blv/10^8 \sin \omega t$.

The dc generator is an ordinary dynamo machine having a multiple-coil winding, the ends of the coils of which are connected to a multiple-segment commutator. The armature is usually rotating, and the field stationary. The field sets up magnetic lines of force, which are cut by the conductors on the revolving armature, giving rise to a generated voltage, which is led through the commutator to a unidirectional external circuit. The iron core is built of laminations of iron insulated from each other by mill scale, or lacquer, or japanning, so that eddy currents which can be generated in the iron core, will be a minimum. The field windings are usually stationary, and the armature rotating; hence low voltage is the usual condition of use of dc generators.

1,200 revolutions per minute. The maximum speed of 3,600 revolutions per minute has been increasingly adopted even for very large machines because high speed means decreased size and weight for a given kilowatt rating and better turbine performance. However, water-wheels and water turbines show best characteristics at much lower speeds, roughly a range of 100 to 600 revolutions per minute. Sixty cycles is the prevailing frequency in the United States. Because of weight and space limitations, 400 cycles is popular in the aircraft industry. Fifty cycles is a common frequency in Europe. See also *Alternator*.

Direct-current generators are built with their de magnetic poles on the stator. Armature conductors on the rotor have ac voltages induced in them as they are rotated; the same principle of flux cutting holds as previously described. An automatic mechanical switching device, called a commutator, is placed on the shaft. It carries fixed brushes, and with its many insulated copper bars connected to the armature coils, it inverts every other alternation of the voltage to give unidirectional, or direct-current voltage at the two armature terminals. It is the commutator that requires the rotor to be the armature so that coils and their switching arrangement always move exactly together. Direct-current generators are generally limited to several thousand kilowatts and their application lies mainly in industrial plants of a specialized nature.

Rotating electric generators are driven by a variety of turbines, notably steam, gas, and hydraulic turbines; by gasoline, gas, and diesel engines. These drivers are described under their appropriate alphabetical entries in this volume. The type of driver used is predominately influenced by the type of fuel economically available at a given site as well as local and regional environmental regulations.

Among the major developments within the last few decades are the superconducting turbine generator and the magnetohydrodynamic generator. The application of superconductivity to synchronous machines with rotating field winding was pioneered at the Massachusetts Institute of Technology, commencing in 1967. The development of large superconducting generators for electric utility applications addresses a number of specific needs of the utility industry. By replacing the conventional copper conductor field winding in the rotor of a synchronous generator with a high-capacity superconducting winding that virtually has zero resistance at cryogenic temperatures, important benefits are gained. The most obvious advantage is the elimination of rotor I^2R loss. The resulting reduction in excitation power is accompanied by a similar reduction in rotor ventilation power requirements. Increased power density and the resulting elimination of stator iron at the armature winding are some of the more subtle, but nonetheless beneficial characteristics of the superconducting synchronous machine. See entry on **Superconductivity.**

The magnetohydrodynamic generator is one in which a thermally ionized gas is forced at high temperature, pressure, and velocity through a duct situated in a transverse magnetic field. An induced voltage appears in the third mutually perpendicular direction (Hall effect), and this voltage may be tapped by electrodes within the duct. If the exhaust gas from the magnetohydrodynamic generator is used to heat steam for a conventional generator, a larger portion of the thermal spectrum can be utilized and the system efficiency may be raised from the tradition value (about 40%) to possibly 50 or 55%. The art of the magnetohydrodynamic generator is not in the high development stage as is the case of the superconducting generator, but active research goes forward. Considerable attention has been given to the potential use of magnetohydrodynamic generator units as infrequently required peaking units. See separate entry on **Magnetohydrodynamic Generator.**

Electric Power Transmission

Energy may be transmitted electrically in overhead wires or underground cables as an electronic flow under pressure, the flow being measured in amperes. The voltage is the electric pressure. Electrical energy is proportional to the product of these two quantities. This energy for all practical purposes cannot be transmitted without some losses, the principal one being a resistance loss, which depends upon the current flow and the size of the conductor. In transmitting a given amount of energy, this loss may be reduced by increasing the voltage, with a corresponding decrease in current. Thus, the use of higher voltages increases the efficiency of electric power transmission by decreasing the conductor loss. In addition, there are economic motivations for increasing transmission line voltages. The power that can be transmitted by a line increases with the square of the voltage and decreases directly with the distance. Therefore, for a fixed distance, a doubling of the previous line voltage will permit about four times as much energy transfer. Over the years, the voltages of new transmission overlays have increased progressively.

For several years, AC high-voltage transmission lines have trended toward higher voltages. The lines range from 115 kV to 765 kV upward. The majority of lines installed in the mid 1980s were at voltages greater than 345 kV. A small part of the 350,000+ miles (565,000 km+) of installed line is DC (up to ±400 kV). To meet electric power requirements, the total is expected to be well over 450,000 miles (725,000 km) by the year 2000. Present accelerated research into superconductivity, which feasibly may impact on electric power transmission as it already has done in connection with generators, is indeed difficult to forecast.

Expansion of transmission lines arises from a number of needs, including (1) to supply load centers from new generating facilities; (2) to provide more interconnections between individual electric utility systems and pools in an effort to improve the reliability of power supply systems; and (3) to better distribute available generating capacity—in effect, in some instances, delaying need for new capacity for a while. The essential economy or savings derived from ties between power systems is that peak loads can be handled with less plant investment than if each region were to install the required capacity to meet its own peaks.

From the inception of transmission, the increase in voltage was based on technology and economics. Early changes in transmission voltage reveal a doubling of the previous level when a change was made. One exception is the single, unique use of 287 kilovolts (kV) from Hoover Dam to Los Angeles. Generally, transmission remained at a level of 230 kV from 1922 until 1953 when the first EHV, 345 kV transmission lines were placed into service. During the mid-1060s, 500-kV-class transmission came into being, with 765 kV coming onto the scene in the early 1970s. See Fig. 13.

The majority of the EHV lines through the 1980s utilized alternating current. However, because reactive compensation for extra-long high-voltage AC lines is expensive and line losses are very high. HVDC has become an economical alternative. Line losses for DC are approximately 33% lower since DC power, by its nature, is not subject to reactive power losses. The construction of a HVDC transmission line is lower in cost, requiring only two conductors per line as opposed to three per line used in AC three-phase transmission. This allows for lighter towers to be used and consequently narrower rights of way. Less insulation is normally needed on a DC line to deliver the same amount of power. The first attempt at DC transmission in the United States occurred in 1970 when an 800 kV (±400 kV) line known as the Pacific Northwest Southwest Intertie was placed into operation. This bipolar, overhead line permitted an exchange of power at peak load times between two regions 850 miles (1368 kilometers) apart.

Terminal conversion equipment, however, continues to be the major obstacle to widespread use of HVDC. Since a terminal is made up of an AC switchyard plus a valve hall and harmonic filters, size and area requirements are excessive. The development of a solid state thyristor valve, however, enabled the cost of HVDC to remain stable, in addition to simplifying and improving the reliability of the conversion process. See Figs. 14 and 15.

As previously mentioned, the energy loss caused by current flowing through the line resistance is not the only loss of energy in power transmission. The long stretches of parallel conductors have a capacitive effect, causing them to draw a current much as a condenser, even though the switches at the far end of the line are open. Furthermore, at high voltages, the air surrounding the conductors becomes partially ionized, and there exists a brush or corona discharge which represents a leakage of energy. The latter, as a matter of fact, is a limiting factor in the raising of electrical pressure on the transmission line and on the size of conductor, and tends to offset the savings due to the reduction in current made possible by an increase in voltage. Good transmission line design requires proper coordination of voltage, wire size, and line losses, so that the desired power may be transmitted at a minimum total annual cost.

When current flows on a long tannsmission line, the inductance and resistance of the line, plus the shunt capacitive effect, cause the voltage at the receiving end to vary with the load, even though the voltage is constant at the sending end. This effect is known as transmission line regulation, and can be calculated quite accurately. From the standpoint of the equipment served, it is desirable that this line regulation be controlled so that utilization equipment eventually supplied from step-down transformers may operate within a reasonable voltage range. Regulators and tap-changing transformers are among the means employed to control transmission line regulation. Furthermore, it can be demonstrated that the economy of power transmission varies as the square of the power factor, making it extremely desirable to operate the transmission line at unity power factor. On account of the induction

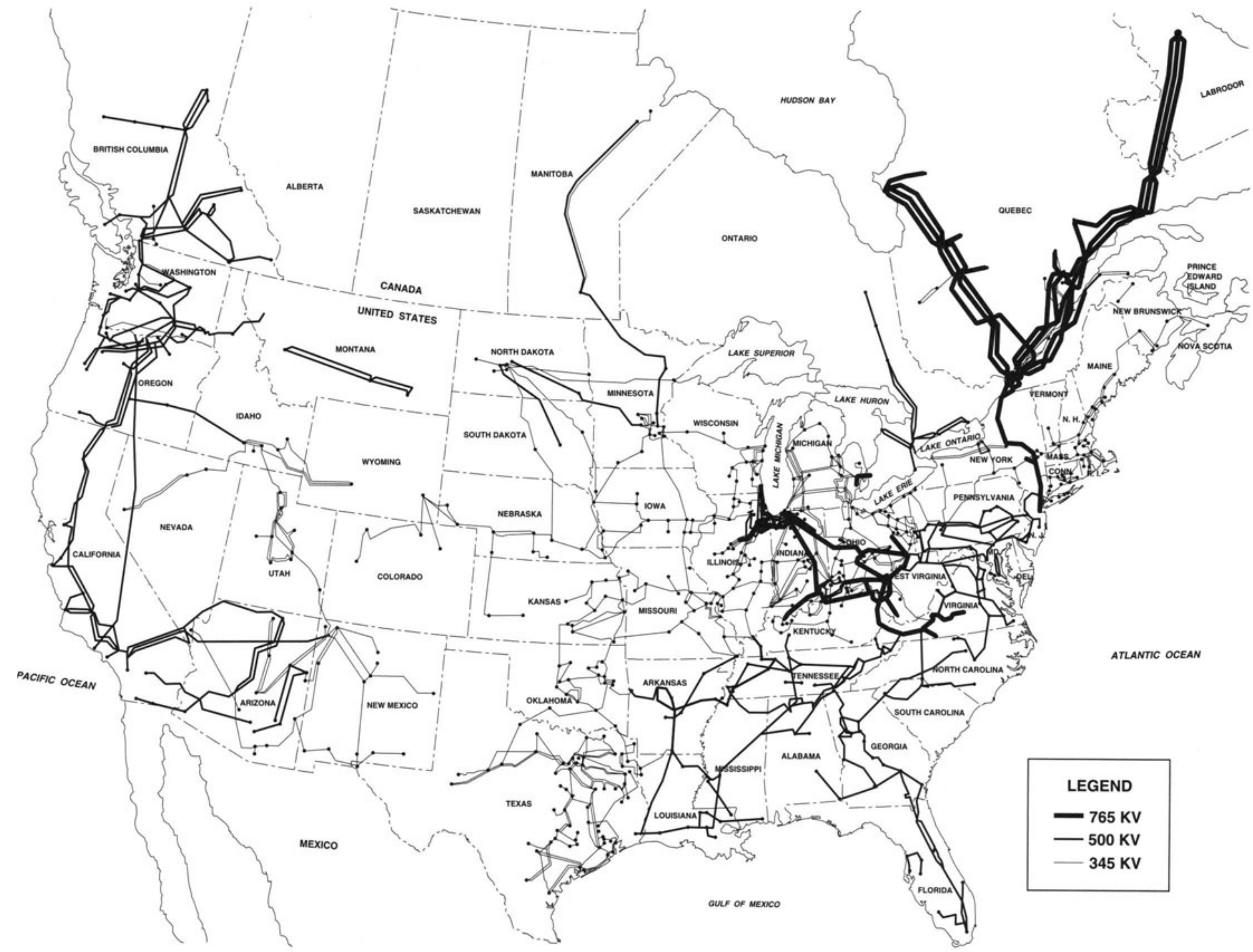


Fig. 13. Transmission networks (extreme-high voltage) and their principal transmission facilities in the United States as of the late 1980s. (*American Electric Power Service Corporation.*)

characteristics of most electrical loads, and the capacitive effect of a transmission line itself, the current on a transmission line will tend to vary from lagging to leading or vice versa with load, although customarily it tends to be lagging. For this reason, a synchronous motor has the advantage of providing some power factor correction. Another method to improve the power factor is to connect static capacitors to the line.

A number of methods have been proposed in past years for VAR (volt-ampere reactance) optimization, but serious shortcomings have prevented them from fully exploiting the VAR and voltage control capabilities of currently available devices. See Fig. 16.

The physical exposure of the ordinary transmission line makes it a likely victim of lightning, and no extensive transmission line could be successfully operated without adequate lightning protection. This may consist of lightning arrestors strategically placed, or an overhead grounded guard wire. Direct strokes are usually immediately dissipated by flashover on the insulators, the lightning then finding its way down the pole to the ground. High-frequency induced waves (2000 to 5000 cycles) may not possess flashover potential, and will travel along the line until discharged by an arrestor, or attenuated, i.e., dissipated in resistance loss. Aside from lightning, a line should be protected against overload and short-circuits. This is commonly the function of transformer fuses, substation circuit breakers, or circuit breakers in general.

In the early 1980s, EPRI established a High Voltage Transmission Research facility, staffed by experts in transmission line research and testing. Located on a large site in Lenox, Massachusetts (near Pittsfield), the facility is well suited to assess the effects of a wide range of weather conditions on high-voltage lines. Surrounding hills shelter the site from high winds and severe vibration, but the facility experiences ample fog, rain, snow, and temperature variations appropriate for electrical and environmental studies. Work at the facility centers on evaluation of insulator performance, corona phenomena, and electric and magnetic fields. Some of the specific studies have included: (1) assessment of the nonbiological field effects of induced voltages and currents on conducting bodies; (2) investigation of ways to reduce radio noise and audible noise emanating from high-voltage lines; (3) evaluation of the electrical strength of insulators subjected to surface contamination; (4) determination of the effect of tower geometry on air gap insulation strength; (5) determination of the effects of switching surge amplitudes on line insulation design; and (6) development of guidelines for hybrid ac/dc lines. See Fig. 17. See also article on **Lightning.**

Turning from electrical to mechanical characteristics of electric power transmission, the material used as a conductor is generally either copper or aluminum, frequently with a steel core for mechanical strength. The former is used in many cases, but the latter is in use for the very high-voltage lines because the larger diameter is effective in reducing corona loss. It also has low relative cost and high strength-to-weight ratio. Except for very low-voltage distribution network lines, the wires are not insulated, and are carried, mechanically held in position, by insulators of porcelain or glass. Sometimes these insulators are mounted rigidly on the cross arm of the poles, but for the higher voltages suspension insulators are used. On high-voltage lines, each insulator is a chain of separate units, so that the voltage from the line to the pole or tower has a uniform gradient across the insulator.

Fig. 14. In the late 1980s, electric utility industry researchers reported that increasing use of solid-state switching systems is being made in phase control and high-voltage dc converters. These systems are composed of a multitude of thyristor (silicon-controlled rectifier) switches connected in series/parallel configurations to provide high-voltage, high-current capabilities. Recently, it has become feasible to trigger high-power thyristor switches by the direct action of infrared (IR) light on a photothyristor. Systems using such thyristors have the advantage that the triggering light signal is carried by electrically insulated fiber optics, and special insulation is not required for the gate circuit. An additional advantage is that the fiber-optic cable is immune to electrical noise pickup and the attendant possibility of destructive accidental triggering. (*Electric Power Research Institute.*)

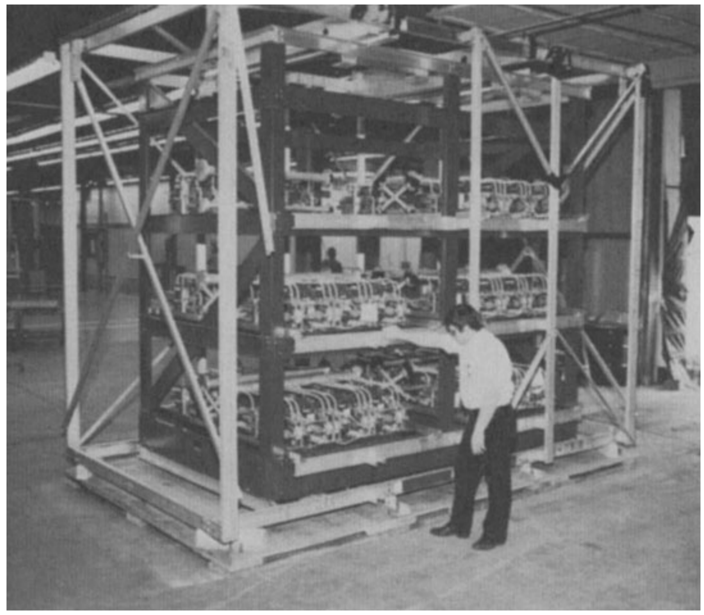


Fig. 15. Since thyristor valves presently account for approximately one-third of the cost of an HVDC terminal, the objective of research has been to develop an advanced thyristor valve that will exhibit improved performance, smaller size, lower losses, and lower cost than traditional thyristor valves. Research targets have included a light-fired, 77-mm-diameter cell, with 5000 V blocking voltage and a low thermal resistance package. Large increases in effective light output intensity have been obtained by advances in the cesium arc lamp design and by using more efficient fiber-optic cables. The new cesium amalgam lamps contain mercury in addition to cesium, and have a reduced bore to increase the arc intensity. As a result of these improvements in the lamp and in the fiber-optic cables, ample light intensity has been obtained at a thyristor gate over an optical path of some 18 meters (59 ft) at moderate pulsing current. For example, 33 times the threshold light intensity has been delivered to an experimental photothyristor by using pulse currents of only 300 amperes. (*Electric Power Research Institute.*)

Galvanized steel, self-supporting towers or poles, or wood H- or K-frame structures traditionally have been used to support overhead transmission conductors. Guyed towers or poles have the least weight, which can be further reduced by the use of aluminum instead of steel. For EHV applications, self-supporting steel towers are most commonly used. The lower voltage substransmission lines are generally supported by wooden poles using cross arms. A commonly used pole is of southern yellow pine, well creosoted. The towers or poles should be sufficiently high so that, at the middle of the span between them, the bottom of the wire sag has a minimum clearance over the ground, as specified in the safety code.

The spacing of poles must be chosen with due regard for temperature and sag conditions. Larger sags in the wire between poles give lower stresses in the wire, and permit the use of longer spans, but at the expense of the use of taller poles to maintain the minimum clearance. Consequently, the observer may see extremes of design representing individual designers' viewpoints, varying from relatively low, closely spaced poles, with little sag, to long spans where tall poles are spanned by wire having a much greater sag. At the same time, the effects of contraction caused by lowering of temperature in winter, and the loads suffered during storms, particularly sleet storms, must be guarded against.

In the early 1980s, EPRI established a Tramsmission Line Mechanical Research Facility (TLMRF), which is located near Fort Worth, Texas. The TLMRF has been acclaimed as the world's most advanced facility for transmission systems structural testing and research. As shown in Fig. 18, the facility is equipped to test towers up to 180 ft (55 m) high by means of three longitudinal and two transverse reaction frames. These frames can apply maximum pulls from any level; maxi-

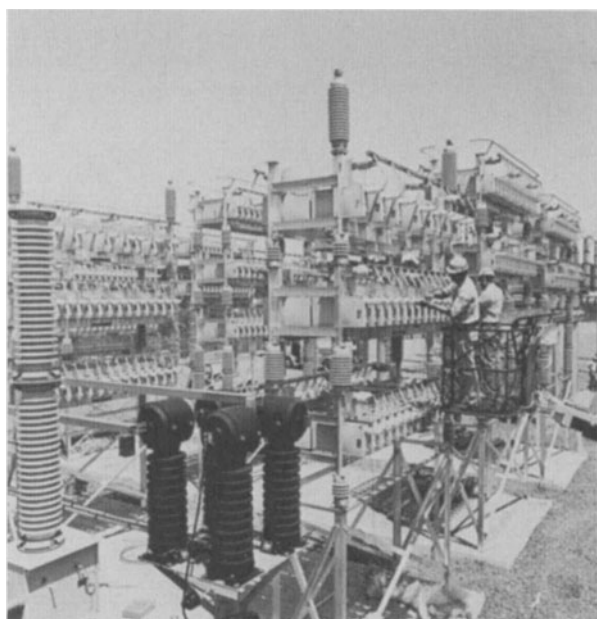


Fig. 16. In the late 1980s, EPRI initiated a project to research methods for improving VAR optimization. The project resulted in a decoupled optimal load flow program suitable for large-scale system analysis. The program has been successfully tested for active power optimization, and the testing of reactive power optimization is in progress. (*Electric Power Research Institute.*)

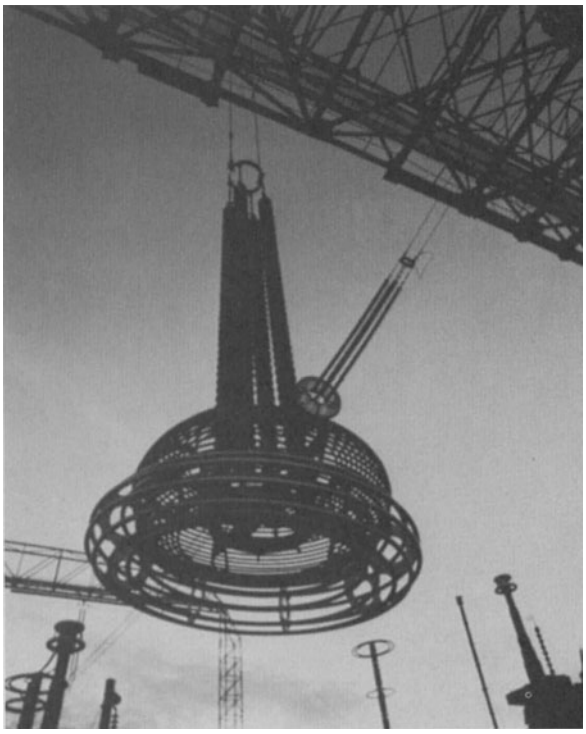


Fig. 17. A portion of EPRI's research and test facility located in Lenox, Massachusetts. The HVTRF (High Voltage Transmission Research Facility) specializes in research on insulators, corona phenomena, electric and magnetic fields, and all aspects of transmission line design and performance. (*Electric Power Research Institute.*)

mum pulls for all transverse loads, and longitudinal loads at one-half the transverse load for each longitudinal reaction structure. Concrete test pads can support all types of transmission structures, up to and including 1200-kV towers. A unique feature of the facility is that both research testing and proof testing can be performed at the same time. A nonenergized 345-kV prototype transmission line is a part of the facility, thus permiting experiments with respect to total structural system behavior under real-life conditions.

Growth in Circuit- and Gigawatt-miles. In order to determine transmission system growth patterns, the parameter "total circuit miles" does not account for the disparity that exists in load capability for each of the voltage classes considered. A more meaningful way to measure this growth is in terms of capability of the transmission system to transmit power. Thus, the term gigawatt-mile is an excellent means of expression since it constitutes a measure of transmission capability. The term is analogous to "passenger miles" as used by the airlines, for example, to express the capability to transport people, or of "ton miles" in terms of freight.

Typical load-carrying capacity of circuits operated within the various voltage classes are shown in Fig. 19.

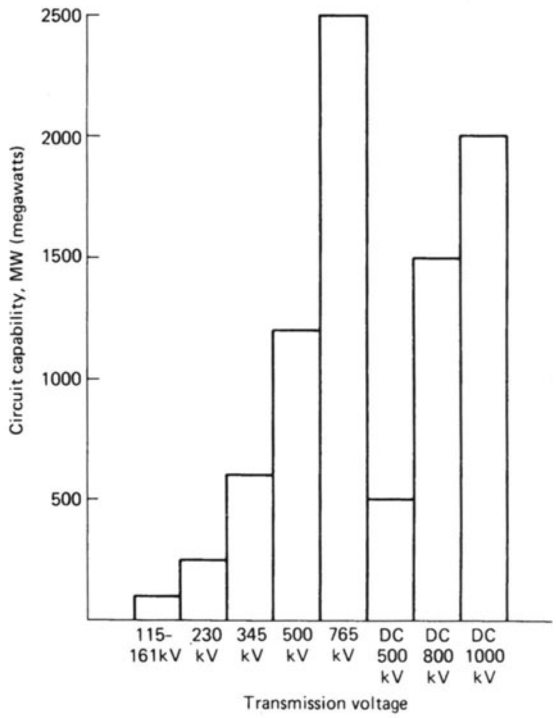


Fig. 19. Typical circuit capabilities.

The impact of EHV on the growth pattern of transmission capability is evident from this chart. Until 1960, EHV accounted for less than 10% of the total capacity, while the lower voltage classes were the real workhorses. EHV growth began in 1964, and by 1970 accounted for three-fourths of the doubling in transmission capability that occurred during that period. Capital expenditure postponements by the utility industry on line construction were responsible for the decline shown in the mid-1970s. Although annual additions for 115-230 kV are expected to decline over the next 10 years, EHV is expected to grow at a greater rate for the same period. Future DC capacity is projected to be minimal for both 800 kV and 1000 kV voltage levels.

Trends in Overhead Transmission. While there are economic benefits for three-phase transmission at higher voltages, economic advantages for shifting from very high voltage three-phase transmission to lower-voltage high-phase-order transmission has also been demonstrated. Both six- and twelve-phase systems are being considered developmentally. A prototype of a six-phase line has been built and is currently being tested to obtain data on radio noise and electric field effects. Mechanical testing will be done under various conditions to observe conductor movement. A significant problem to be solved in multiphase transmission (6 or more phases) is relaying and protection.

Reduced rights-of-way and the desire to reduce transmission losses are among the factors that have forced the utility industry to higher-voltage transmission systems. The development of technology for 1200 kV lines and equipment has been in progress in the United States since 1967. Among the problems of ultra-high voltage transmission are radio and audible noise, induction effects, corona and generation of ozone, and mechanical stresses imposed upon the multiple-conductor bundles used. Studies have shown benefits for a single UHV line, compared to two 500 kV lines, when comparing line constructing costs, annual cost of delivered power, and cost of losses. Reliability of a single UHV line is less, however, than two EHV lines, since the single-line probability of outage is greater than that of two lines.

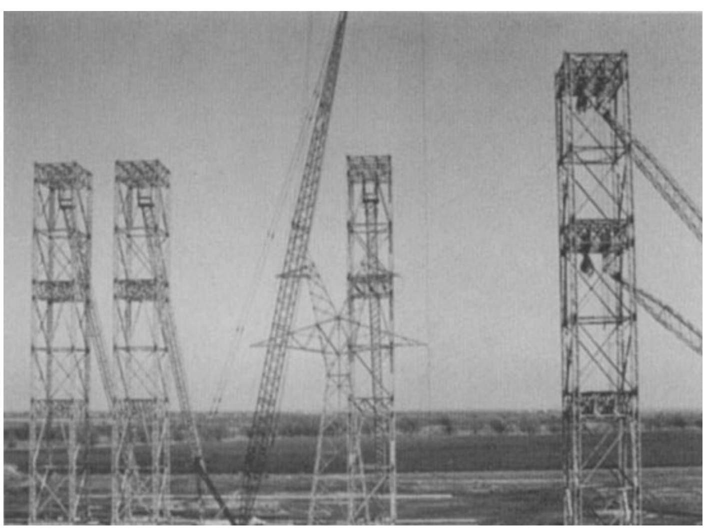


Fig. 18. View of transmission tower testing frames in place at the EPRI Transmission Line Mechanical Research Facility (TLMRF), located near Fort Worth, Texas. (*Electric Power Research Institute.*)

Considerable research and development efforts are underway in the area of DC transmission with concentration on developing more compact, lower-cost terminal converter facilities, increased voltage levels, and the expansion of the market to multi-point projects. Specifically, funded projects include HVDC transmission line research, thyristor valve improvement and uprating, HVDC circuit breaker development, and HVDC/HVAC interaction and system studies.

Underground Transmission Technology. Overhead transmission systems are the most logical choice for utilities unless the overhead lines are barred by topography, legislative action, or nonexistent or extremely expensive rights-of-way. Cost has now become a strong factor to consider when planning transmission expansion. The generally accepted ratio of installed costs, underground to overhead, that ranged between 10:1 and 20:1 is no longer used. Studies show that no general comparison between system costs is practical due to the different attributes of each new circuit. The various differences include price and availability of land, population distribution, terrain, load size, circuit length, and local regulations. Local input is becoming an ever larger influence in the transmission planning process.

Although comparisons of the two technologies for rural installations show cost ratios as low as 5:1, cost is still a heavy disadvantage of going underground. In a typical city installation, for example, where the majority of underground circuits are employed, nearly 50% of the circuit cost is incurred in cutting the trench, pulling the cable, backfilling the trench, and reconstructing the surface. Consequently, only 1% of all transmission circuit-miles in service in the United States is installed underground. For this percentage to change, new and improved methods of underground transmission will have to be developed. Research is being sponsored for the development of new cables and more efficient underground installation methods that will result in decreased cost and more reliable systems.

Figure 20 summarizes the new and existing types of underground transmission systems with a time schedule showing their dates of commercial application. A brief discussion of each of these technologies follows.

Pipe. Over 75% of the underground circuits in the United States consist of pipe-type cable systems. This type of system utilizes three oil-impregnated paper-insulated cables enclosed in a single steel pipe with oil under high pressure. This high-pressure, oil-filled (HPOF) cable has been in service since the mid-1970s and designs rated up to 500 kV have been successfully tested. Limitations that still exist with HPOF cables include excessive overload temperature rise, high losses in the insulation, and severe charging currents at higher voltages. In addition, they are subject to occasional failures. High-quality semi-synthetic insulation systems and forced-cooling techniques are being investigated, as are rapid techniques for fault location.

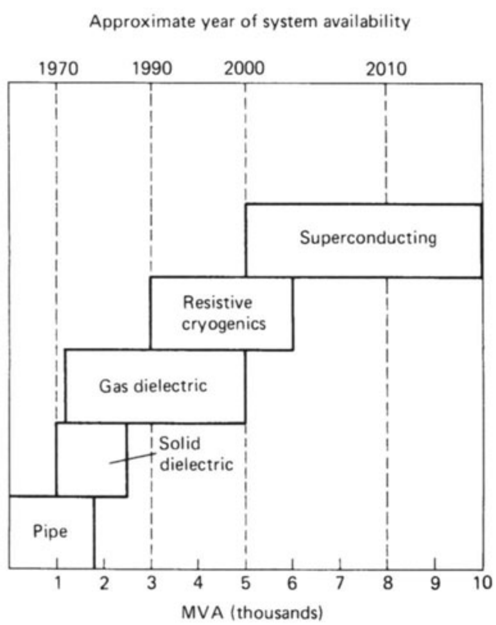


Fig. 20. Comparative capabilities of underground transmission systems.

Gas Insulated. Both rigid and flexible gas-insulated cables have been developed. In this system, sulfur hexafluoride gas is used for the insulating and cooling medium. Rigid cables of a single-phase design, while capable of transmitting large quantities of power, are only economical for short runs. To overcome this problem, researchers developed a three-conductor, gas-insulated cable. Overall system size is reduced and a 15–25% reduction in cost is possible. Installation is easier and more efficient, because splicing and field welding are reduced considerably. See Fig. 21.

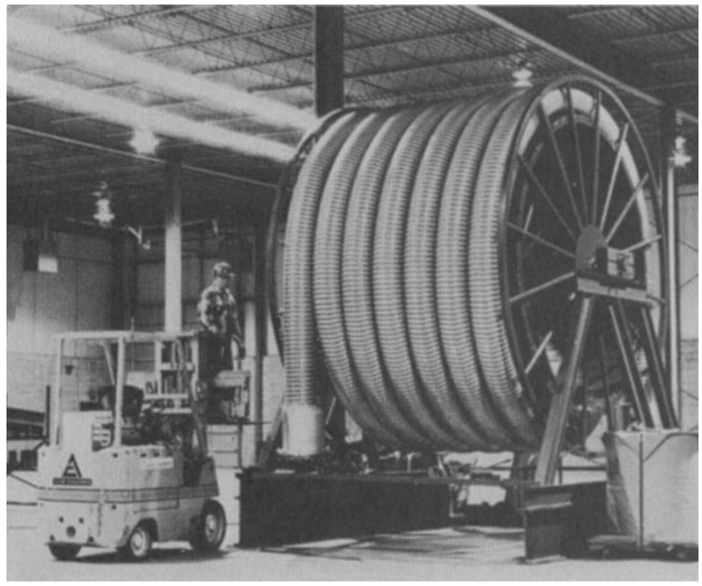


Fig. 21. Flexible gas-insulated transmission line under test at an EPRI facility. The development of flexible gas cable extends the advantage of SF_6-insulated cable (low losses, inexpensive termination, and no oil handling equipment) to lower power levels (from 2000 MVA to 1000 MVA at 345 kV) and considerably improves the economics and ease of installation. It has been estimated that the flexible system will be about 25% less expensive on an installed basis as compared with the rigid-isolated phase gas cable. (*Electric Power Research Institute*.)

Solid Dielectric. This system normally utilizes cable of the extruded polyethylene type, where a solid insulation of the synthetic material surrounds the conductor. These cables are easier to handle than the previous types because no insulating fluid is involved. The reliability of these cables, however, is heavily dependent on the purity of the dielectric. Any contaminants can cause electrical discharges that result in failure. Methods have been developed, and are now being perfected, to combat these contaminants and enhance cable reliability. Development and testing is currently being carried out for cable designs at 138 kV. A 230 kV cable has also been produced and will be tested after splice and terminal developments have been completed. A 345 kV cable is under development.

Resistive Cryogenics. Since electrical resistance is a function of the conductor temperature, it is possible to reduce the heat losses in a system if operation is at low temperature. The two conductor metals, copper and aluminum, have electrical resistances at liquid nitrogen temperatures (−320°F; −196°C) of one-tenth the magnitudes at normal ambient temperatures. In order for cryogenic systems to be considered efficient, reductions in resistance have to generate savings that match or exceed the cost of refrigeration. Cryogenic systems are currently limited by the high cost and low efficiency of gas expansion-compression refrigerators. Recent attempts to develop a magnetic refrigerator have fallen short of both design goals and what would be needed for a practical refrigerator.

Superconducting. Superconducting systems operate at even lower temperatures, down into the liquid-helium range of −452°F (−266.9°C).

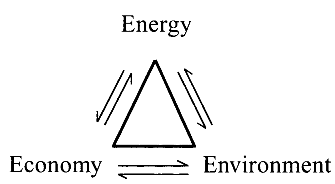
Certain metals, such as niobium, lose their electrical resistance at these temperatures with the result that extremely high currents can be transmitted with low losses. Superconducting systems have demonstrated overall efficiency, that is, there is a net gain when savings and costs are compared. Research on superconductivity is one of the most aggressive areas in both physics and chemistry as of the early 1990s. See **Superconductivity.**

Basis for Electric Power Research

There are several goals that science and advanced technology can target for solving global problems and ultimately achieving a better, more meaningful life for all peoples. Solutions for many problems obviously remain, but within the past decade in particular, technology has accelerated and numerous answers (what might be termed *intermediate* answers) have resulted. For example, science has contributed effectively toward identifying problems and the general application of the principles of physics and chemistry to the economic and sociological aspects of energy-related problems. Further, a lot of background evidence has been collected, much of which has been derived from sophisticated measurement methods. This has made industry and the general populace aware of the need for advanced nations to tackle what seem to be almost innumerable problems, if people are to continue a safe and rewarding life on Earth and, further, to shield future generations from a deteriorating planet.

The triangle below (repeated from the entry on **Coal Conversion (Clean Coal) Processes**) illustrates the inevitable conflicts between three great factors—the economy, the energy, and the environment. These conflicting forces will remain until drastically new progress is made in our *energy supply* situation.

Economy. The economy of developed nations and of underdeveloped nations (in particular) depends on the availability of reasonably



priced energy. Without affordable energy, the impact of energy costs on a national economy can be tremendous. The perfection of new, environmentally attractive energy sources can be achieved only at the expense of the economy. (There may be an exception in the case of nuclear energy, but this topic enters into the realm of public fears and apparently—with the exception of a few nations, such as France—is not politically viable in the absence of an effective educational program.) See also article on **Nuclear Power**.

Environment. The triangle also shows the inevitable conflict between energy and the environment. See also **Acid Rain; Pollution (Air);** and **Water Pollution.** The environment also conflicts with the economy because the effective partial or intermediate environmental solutions achieved thus far for traditional fossil fuels carry a great expense.

Environmental control equipment can account for 40% of the costs of a new power plant. This does not include operation and maintenance of the equipment, which amounts to perhaps 25–30% of operating costs. See Caruana reference listed.

Energy. The impact of energy costs on the economy have been demonstrated dramatically, notably in the early 1970s at the time of the Middle East oil embargo. Although a large anti-coal sentiment has developed, particularly in the United States during recent years, the fact remains that coal represents a very significant portion of the nation's potential energy supply. Experience has demonstrated that technology also has its own problems (i.e., economic investment alone cannot hasten technical development beyond a certain pace). Technology has an inertial factor, not always appreciated by scientists and laypeople alike. Breakthroughs cannot be legislated. Consequently, from a very practical standpoint, in the absence of a political and sociological solution to nuclear energy, it appears that, to remain a leading nation of the world, the United States cannot turn away from coal as a significant energy source. Logic thus dictates that using clean coal processes, including greatly improved combustion schemes, is of high priorty for advanced technology. Thus, for those intermediate solutions previously mentioned, much funding and effort is being directed toward reducing and possibly ultimately eliminating serious damage to the environment.

Less frequently stressed is how the design of electricity-consuming equipment and conservation efforts on the part of consumers can reduce energy use and thus lessen the environmental impact. Although energy waste often is obvious, it is left uncontrolled. As one example, witness the overcooled buildings in summer; the overheated offices, apartments, and houses in winter!

Electric Power Technology Research Needs and Goals

Few major commercial and industrial enterprises interact with and depend upon so many segments of science and technology as the electric power industry does—in both the United States and other advanced industrial nations. Just analyzing and classifying the many research projects currently underway and to be continued and projecting research needs far into the future is a massive undertaking. A laudable effort along these lines began over a decade ago by a consortium of electric power suppliers across the country and in some cases in collaboration with electricity suppliers in other parts of the world.*

I. Extension of Electric Power Applications

1. Develop new appliances, tools, processes, manufacturing methods, and so on that can use electricity more efficiently or that substitute electric power for other energy sources that presently are pollutive and inefficient.
2. Develop improved electric power delivery systems that are better tailored for specific end users.
3. Develop advanced refrigeration systems that do not require the use of polluting chlorofluorocarbons.
4. Develop new concepts in electric transportation systems that do not rely on polluting fuels. This includes research on electric cars, more efficient motors, batteries, and solar power.

II. Environmental, Health, Welfare, and Safety Issues

1. Electric and magnetic fields (EMF) assessment. Technologically, this remains a poorly understood subject as of 1993. However, the electric power industry must continue to evaluate all new findings and conduct research on possible health risks and, should the problem be assessed as a health risk, develop options and strategies to mitigate such proven risks.
2. Provide scientific and technical basis, as well as risk management tools for electric power industry response to the climatic change issue. (This concerns the still-controversial issue of the effects of carbon dioxide on global warming.) See **Climate.**
3. Develop remedial measures, as may be required from future studies of the climate. Subprojects include development of an air quality study, air toxics management, and environmental risk management.
4. Improve technological approaches to ground/surface waters protection, solid waste (including radioactive), disposal, and numerous occupational health issues.

III. Sustaining the Future of Electric Power

1. Study the safety/economics of nuclear plants, innovative nuclear concepts, and advanced fossil fuel technology. Subprojects include improvement of combustion turbines, corrosion research to increase operating safety and reduce materials costs, coal gasification/fuel cells, and the concept of distributed smaller power plants versus the large central stations.
2. Study and develop renewable energy resources, including geothermal, solar, biomass, and so on.
3. Explore advanced transmission technology, leading to the development and demonstration of methodologies and strategies that are more flexible and less costly by new approaches, such as by the use of superconductors.
4. Research energy storage systems, including compressed air, battery, and suprconducting energy storage systems.
5. Conduct exploratory research to assure a fundamental knowledge base to support the development of cost-effective and environmentally sound advanced supply, delivery, and end-use technologies.

*"Research & Development Plan" prepared by the Electric Power Research Institute, Palo Alto, California. This publication describes in detail several research projects already completed, underway, or scheduled for future attention, as of 1993.

IV. Electric Power Production Cost Control

The electric power industry currently faces increasing costs at a time when competitive pressures are increasing. Numerous programs are underway for determing the cost and productivity of fossil fuel, hydropower, and nuclear power plants. Among the specific projects are those directed toward better instrumentation and automatic control of plant operations, provision of suburban underground transmission and delivery systems, automated maintenance, and computerized load management systems.

Additional Reading

Abelson, P. H.: "Reliability of Electric Service," *Science*, 689 (August 18, 1989).

Ahner, D. J.: "Benefits of Combined Cycle Cogeneration," *Chem. Eng. Progress*, 49 (March 1988).

Beardsley, T.: "Guessing Game: Electromagnetic Fields," *Sci. Amer*., 30 (March 1991).

Beaty, H. W.: "Electrical Engineering Materials Reference Guide," McGraw-Hill, New York, 1990.

Blatt, M., and M. Khattar: "Water-Loop Heat Pump Enhancements," *EPRI J*., 37 (July/August 1992).

Caruana, C. M.: "EPRI's Research Key to Curbing Utilities Costs, Protecting Environment," *Chem. Eng. Progress*, 72 (November 1987).

Chapman, S. J.: "Electrical Machinery Fundamentals," 2nd Edition, McGraw-Hill, New York, 1991.

Chexai, B.: "CHECWORKS: Integrated Corrosion Software," *EPRI J*., 42 (July/August 1992).

Considine, D. M., Editor: "Energy Technology Handbook," (1857 pp.), McGraw-Hill, New York, 1977. (A classic reference of the post-oil embargo period energy R&D. Out of print.)

Corcoran, E.: "Cleaning Up Coal," *Sci. Amer*., 106 (May 1991).

Crawford, M.: "Electricity Crunch Foreseen…Maybe," *Science*, 1005 (November 18, 1988).

Douglas, J.: "Taking the Measure of Magnetic Fields," *EPRI J*., 16 (April/May 1992).

Douglas, J.: "Advanced Motors Promise Top Performance," *EPRI J*., 24 (June 1992).

Fink, D. J., and H. W. Beaty: "Standard Handbook for Electrical Engineers," 12th Edition, McGraw-Hill, New York, 1987.

Fitzgerald, A. E., and C. Kingsley, Jr.: "Electric Machinery," 5th Edition, McGraw-Hill, New York, 1990.

Forsyth, E. B.: "Energy Loss Mechanisms of Superconductors Used in Alternating-Current Power Transmission Systems," *Science*, 391 (October 21, 1988).

Fri, R.: "The Challenge of Global Sustainability," *EPRI J*., 14 (June 1992).

Frischmuth, R.: "Tools for Gas Turbine Management and Maintenance," *EPRI J*., 38 (June 1992).

Goldstein, R., and D. Porcella: "Biofilm Formation and Microbial Corrosion," *EPRI J*., 34 (July/August 1992).

Gonen, T.: "Electric Power Distribution System Engineering," McGraw-Hill, New York, 1986.

Hertz, N., Stewart, N., and A. Cohn: "High-Efficiency GCC Power Plants," *EPRI J*., ((July/August 1992).

Jaret, P.: "Electricity for Increasing Energy Efficiency," *EPRI J*., 4 (April/May 1992).

Johnson, K.: "Commercial Building Energy Analysis Tools," *EPRI J*., 30 (April/May 1992).

Hunter, L. C., and L. Bryant: "The Transmission of Power," The MIT Press, Cambridge, Massachusetts, 1991.

Kalra, S. P.: "Outage Risk Management (Nuclear Plants)," *EPRI J*., 34 (April/May 1992).

Kesserling, J.: "Microwave Clothes Dryers," *EPRI J*., 34 (June 1992).

Kraus, J. D.: "Electromagnetics," 4th Edition, McGraw-Hill, New York, 1992.

Lowdon, E.: "Practical Transformer Design Handbook," 2nd Edition, McGraw-Hill, New York, 1988.

Moore, T.: "Refrigerants for an Ozone-Safe World," *EPRI J*., 22 (July/August 1992).

Morris, N.: "Mastering Electrical Engineering," 2nd Edition, McGraw-Hill, New York, 1991.

Mueller, F. M.: "Superconductor Landscape: View from Los Alamos," *Advanced Materials & Processes*, 79 (January 1988).

Nye, D. E.: "Electrifying America," The MIT Press, Cambridge, Massachusetts, 1991.

Rassenti, S. J., and V. L. Smith: "Smart Computer-Assisted Markets (Electric Power Networks)," *Science*, 537 (October 25, 1991).

Rastler, D.: "Distributed Generation," *EPRI J*., 28 (April/May 1992).

Rosen, J.: "Prediction and Prevention (Electric Power)," *Technology Review (MIT)*, 19 (April 1991).

Ross, M.: "Improving the Efficiency of Electricity Use in Manufacturing," *Science*, 311 (April 21, 1989).

Schneider, T.: "HTSC Power System Consortium Proposed," *EPRI J*., (April/May 1992).

Seldman, A. H., and H. Mahrous: "Handbook of Electric Power Calculations," McGraw-Hill, New York, 1984.

Staff: "Electricity Transfers and Reliability," North American Electric Reliability Council, Princeton, New Jersey, 1990.

Staff: "Study Claims Electric Utilities Lag in Emission Control," *Chem. Eng. Progress*, 14 (October 1990).

Staff: "EPRI Bolsters Pulp and Paper Presence," *Chem. Eng. Progress*, 15 (November 1990).

Staff: "Academy to Wade into the Electromagnetic Field Quagmire," *Science*, 955 (August 30, 1991).

Staff: "EMF and Male Breast Cancer?" *Science*, 964 (August 30, 1991).

Staff: "Reliability Assessment 1991–2000," North American Electric Reliability Council, Princeton, New Jersey, 1991.

Staff: "EPRI Research & Development Plan," Electric Power Research Institute, Palo Alto, California, 1992.

Staff: "Enhancing Productivity Through A Wider Use of Electricity," *Issues and Responses*, 8, Electric Power Research Institute, Palo Alto, California, 1992.

Staff: "Ensuring A Continuous Electric Power Supply," *Issues and Responses*, 10, Electric Power Research Institute, Palo Alto, California, 1992.

Staff: "Developing New Ways to Protect Air, Land, and Water," *Issues and Responses*, 13, Electric Power Research Institute, Palo Alto, California, 1992.

Staff: "Heat Pumps," *EPRI J*., 10 (April/May 1992).

Staff: "Electric Vehicles," *EPRI J*., 12 (April/May 1992).

Timberman, M.: "Advanced Metering Benefits on Both Sides," *EPRI J*., 18 (April/May 1992).

Weiss, J.: "ACOM: Availability Cost Optimization Methodology (Electric Power Generation)," *EPRI J*., (June 1992).

Wendland, R., and M. Blatt: "Copl Storage Saving Money and Energy," *EPRI J*. 14 (July/August 1992).

White, D. C., Andrews, C. J., and N. W. Stauffer: "The New Team: Electricity Sources Without Carbon Dioxide," *Technology Review (MIT)*, 42 (January 1992). Good chart on amount of money spent on energy R&D.

ELECTRIC POWER RESEARCH INSTITUTE (EPRI). Established in 1972 as an independent nonprofit consortium of U.S. electric utilities, with a current (1992) budget of over $0.5 billion and headquarters located in Palo Alto, California (93404), the mission of EPRI is to discover, develop, and deliver advances in science and technology for the benefit of member utilities (more than 700 in number, representing about 70% of U.S. electric utilities). Because hundreds of utilities pool their research funds in the EPRI program, the Institute is enabled to undertake developments on a scale that no one utility could handle. EPRI sometimes co-funds projects with other R&D organizations, thus further leveraging its members' research investments. Because of the breadth of its work, EPRI has developed strategic relationships with other R&D organizations throughout the world, cooperating in international research endeavors and serving as a clearing house of energy-related information for its members as well as for the scientific community.

EPRI targets four main R&D targets:

1. *Electric power generation,* focusing on developing cleaner, safer, and more economical ways to produce electric power, including, for example, the development of robots for use in maintaining existing fossil and nuclear power plants, and the development of entirely new systems, such as compressed air energy storage and solar photovoltaics.
2. *Electric power delivery,* including efficient and economical ways to transmit, transform, and distribute electricity in harmony with the environment. Products range from computer software that designs and operates the delivery system to procedures that extend the life of delivery equipment.
3. *Energy use*, for advancing the more economical and efficient use of electricity in the home, office, and factory. Developments range from efficient space heating and cooling products, for residential and commercial use, to plasma and infrared processes, for industrial applications, to electric vehicles for urban travel.
4. *Environmental science,* with major programs of scientific research and analysis into health and environmental concerns, such as electromagnetic field effects and global climate change.

The Institute maintains an on-line communications network (EPRINET).

ELECTRIC POWER (Tidal). See **Tidal Energy.**

ELECTRIC RAYS. See **Skates and Rays.**

ELECTRIC RESISTANCE. See **Resistance.**

ELECTRIC SHOCK. Experiments with human volunteers carried out by various researchers since 1933 have established that the mean threshold of perception (the first "tingle") experienced from an electric shock is 1.1 milliamperes (mA) for men and 0.7 mA for women. Slightly higher current (called reaction current) provides for the possibility of an unexpected involuntary reaction which might result, for example, in the dropping of a hot pan of grease or a fall from a ladder. In 1967, the American National Standards Institute (ANSI) and the Underwriters Laboratories (UL) were funded to determine reaction current levels. The result of these studies was an ANSI specification in 1970 which limited maximum current leakage levels to 0.5 mA for new, properly operating, 2-wire cord-connected appliances; and 0.75 mA for heavy, movable cord-connected appliances, such as freezers and air conditioners. While this specification predicts that new appliances should never give perceptible shocks, this may not be true after the hardware has aged for a number of years or has been misapplied or exposed to degrading environments. It should be emphasized that a poorly designed or poorly used 120-volt appliance is capable of delivering lethal shocks.

As the intensity of shock current increases, sensations of warmth, tingling, and muscular reaction increase and pain develops. Eventually, a current level is encountered at which a person cannot voluntarily "let go" because of muscular spasms and thus the victim is "frozen" to the circuit. This threshold is of great importance because under electrically induced muscular spasms, the musculature of breathing becomes paralyzed. Unless the circuit is interrupted, collapse, unconsciousness, and death may follow in a matter of minutes or fractions thereof. Research has shown that this threshold for men is 16 mA and 10.5 mA for women See Fig. 1. In the case of higher-current shocks, resumption of normal breathing may not occur for several minutes after the current is stopped In these cases, artificial respiration is required to prevent physiological damage. At even higher current levels, heart action becomes seriously impaired or stopped, most frequently by the mechanism of ventricular fibrillation (VF).

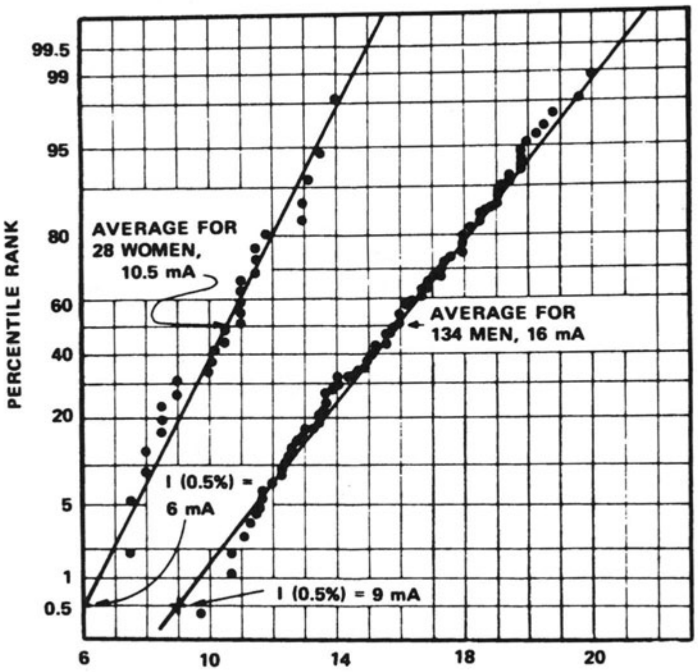


Fig. 1. "Let-go" currents for men follow a normal distribution, as do those for women, using a smaller test sample. Mean values (men and women) are in a ratio of two to three. (*After Dalziel.*)

Ventricular fibrillation, the most common cause of death from electric shock, is the result of desynchronizing the natural electrical impulses which govern the ordered muscular reactions which cause the heart to pump blood. In VF, the heart seems to "quiver" like a bag of jelly without effective pumping. Brain death normally occurs within 4 to 6 minutes when this condition sets in. After the onset of VF, normal heart action seldom resumes spontaneously and unless prompt defibrillation is effected by trained medical personnel, death is inevitable. Defibrillation is accomplished by applying a stronger, shorter duration shock of controlled amplitude and duration to electrodes on the chest. This stops the quivering for an instant and then allows the normal, synchronized heart action to resume. Cardiopulmonary resuscitation technique allows maintenance of circulation by rescuers for many minutes, even for hours with help, until defibrillation can be accomplished. See Fig. 2.

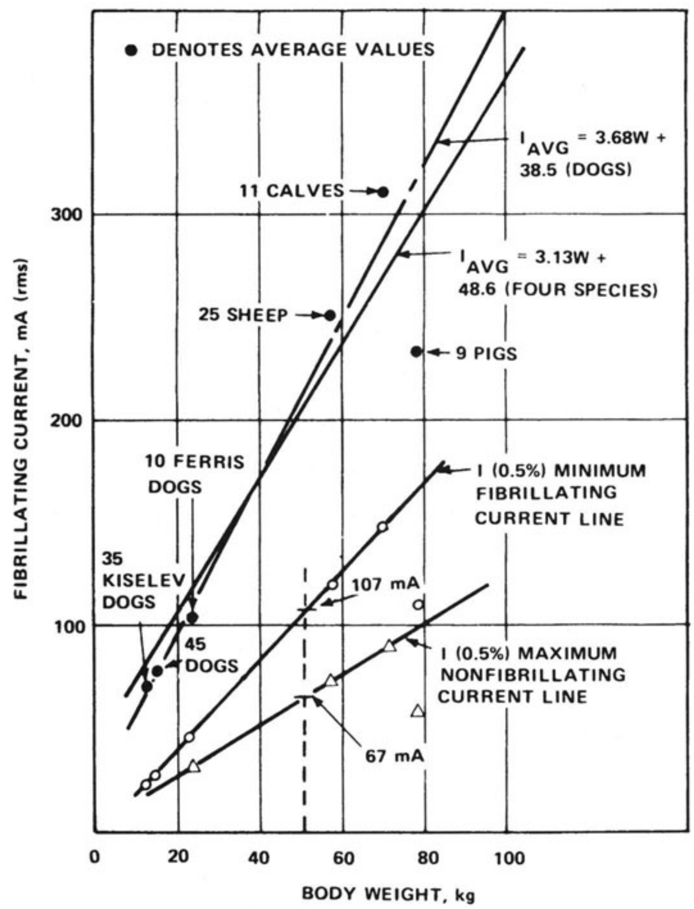


Fig. 2. Fibrillating current versus body weight for animals. These curves combine results of several experiments, including those on calves and pigs. (*GTE Laboratories.*)

The complications that follow an electric shock vary considerably and depend mainly on the amplitude and duration of contact with electric current. In addition to heart complications, electrical burns and injuries caused by falling will require immediate attention. In addition to ventricular fibrillation previously mentioned, cardiopulmonary resuscitation may be required because of (1) tetany of the breathing muscles, which is usually limited to the duration of exposure to the current; if the latter is prolonged, tetany may persist and cardiac arrest may occur because of anoxia; (2) prolonged paralysis of breathing muscles which may result from a massive convulsive phenomenon. This may persist for minutes after the shock current has terminated and may result in anoxic cardiac arrest. Cardiopulmonary resuscitation is only effective when performed on a victim who is in the horizontal position.

Of critical importance, the rescuers must make certain that they are in no danger of electric shock.

Considerable research has indicated that a normal, healthy 50-kilogram (110-pound) adult will not be likely to fibrillate if the current intensity is less than

$$I \text{ (electrocution threshold)} = \frac{116}{\sqrt{T}} \text{ mA}$$

where T is in seconds. The relationship to body weight appears to be direct; a 25-kilogram (55-pound) child's electrocution threshold is believed to be one-half the values given by the aforementioned formula. These values have been used in preparation of UL specification 943 which all UL labeled and listed ground fault circuit interrupters (GFCI) must meet. The National Electrical Code, which provides the basis for most building codes and inspections in the United States, requires the exclusive use of UL approved devices.

The National Electrical Code (NEC) requires the use of GFCIs in all receptacle circuits of swimming pools, bathrooms, garages, and outdoor receptacles for new construction. In addition, all single-phase 120-volt receptacles used at construction sites or for other temporary use are required to have GFCI or specified other safety grounding measures.

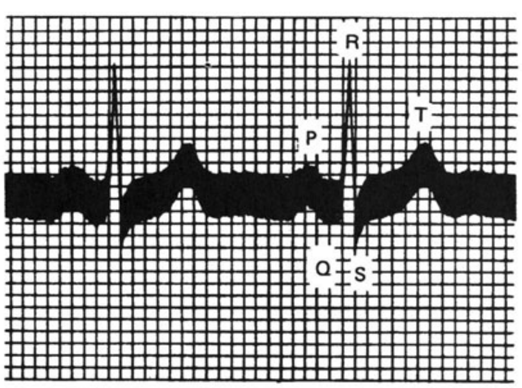
ELECTRIC WELDING. See **Welding.**

ELECTROACOUSTICS. See **Acoustics.**

ELECTROCAPILLARITY. The surface tension between two conducting liquids in contact, such as mercury and a dilute acid, is sensibly altered when an electric current passes across the interface. As a result, when the contact is in a capillary tube, the pressure difference on the opposite sides of the meniscus is affected by a current traversing the capillary column, to an extent dependent upon the direction of the current across the boundary.

ELECTROCARDIOGRAPHY. An electrocardiogram (ECG) is a graphical record of the electric potentials produced by activity of the heart. The normal beating of the heart is associated with the production of bioelectric currents in the organ. These currents are not strong, but they are carried to the surface of the body, where they may be measured by sensitive electrical instruments. The heartbeat normally is controlled by a rhythmically occurring electrical discharge that emanates from a spot in the right atrium known as the sino-artrial node (SA). This wave of electrical activity is spread through the heart muscle by a network of fibers known as Purkinje's network. As the wave spreads, the muscle of the heart rhythmically contracts and then gradually relaxes. The contraction, or squeezing of the muscle forces blood from the heart into the blood vessels, producing the needed pumping action.

As the wave of electrical activity spreads across the surface of the heart, the wave produces an electrical field which spreads through the conducting medium of body tissues to the surface of the body. This



Electrocardiogram of a normal subject.

field can be considered to be produced by an electric-dipole moment which rotates and varies in amplitude in the course of the cardiac cycle. The pattern shown in the diagram was produced by placing electrodes on the left and right arms. The characteristic features of this electrocardiogram are labeled as conventionally described by physicians.

The key wave is due to electrical activation of the atria of the heart; the QRS complex occurs at the time that the ventricles (or main pumps) of the heart contract; the T wave is caused by the repolarization or resetting of the electrical system following contraction. The contracting phase of the heart cycle is called *systole*, and the relaxed phase is known as *diastole*. Changes from the normal electrocardiographic pattern can be used to diagnose a heart attack and the progress and recovery from it. The pattern also can be used to diagnose and follow irregular heart rates (*arrhythmias*) and conditions of faulty conduction.

At the surface of the body, the voltage amplitude of a typical QRS complex will vary between 100 microvolts and 2 millivolts. Faithful reproduction of an electrocardiogram can be achieved with a frequency response flat between 0.25 and 100 Hz. There are some pathological situations which produce "splintering" of the QRS complex that can be easily detected only with a frequency response to approximately 1,000 Hz. This is not a common requirement and most recorders do not have a response above 100 Hz. To produce a faithful electrocardiogram, the voltage-amplifier input impedance should be above 0.1 million ohms and preferably above 1 million ohms.

In order to avoid motion artifacts in the record, care must be taken in the choice and application of the electrodes. Chemical action will cause a gradual change in the impedance of the skin-metal interface which slowly will reduce the amplitude of the ECG and, more seriously, will produce an abrupt change in the dc level, as this resistance changes with relative motion between the skin and the electrode. If silver is used as the electrode, chloride ions from the sodium chloride in body perspiration gradually will deposit on the silver reducing the effective surface area and thus increasing electrode resistance. By coating the silver with a silver chloride surface and by further introducing a paste salt bridge between this surface and the skin, the process can be reduced by several orders of magnitude. The skin has a high resistance compared with the subdermal tissues. Thus, for good results, the skin usually is rubbed briskly before application of the electrode paste and the electrodes. Skin resistance typically can vary from 5,000 to 200,000 ohms under varying conditions of perspiration and surface condition—hence the need for a high-impedance input.

Since the body usually is in a 60-Hz electric field and the room in which an electrocardiogram is made is not shielded, the 60-Hz pickup between two points on the body frequently will be one or two orders of magnitude greater than the desired ECG signal. To reduce this effect, ECG amplifiers employ several differential-input stages and a third electrode is attached to an arbitrary point on the body. This latter electrode is used as a "ground" or common-mode reference point. An ECG amplifier will have a common-mode rejection ratio of approximately 3,000:1.

Vector Cardiography. A variant on the conventional electrocardiographic presentation attempts to show the locus of the basic dipole moment which is producing the electrocardiogram. For this purpose, instead of using the usual pair of electrocardiographic electrodes, a minimum of two pairs of electrodes are placed on the patient's thorax so that the vector sum of the instantaneous voltages appearing across each pair will show at any instant the direction and magnitude of the voltage emanating from the heart. The two output signals thus derived are passed through appropriate weighting networks and then displayed on an oscilloscope, presenting the locus of the tip of the potential vector as it rotates with time through the cardiac cycle as a closed loop.

Processing Electrocardiograms. Because of the importance of the electrocardiogram as a diagnostic tool and the skill required in interpreting it, means have been developed to increase the efficiency of interpretation. Telephone links can be used which permit taking an electrocardiogram by a general practitioner or at a small hospital and transmitting the information to a specialist in facsimile form for immediate analysis and telephonic diagnosis. A frequency-modulated audio tone generator connected to the transmitting ECG machine can be used. At the receiving telephone, the signal is demodulated and fed into a conventional ECG recorder. Monitors are available which permit the patient to wear a small, battery-operated electrocardiograph which telemeters by FM radio over a relatively short distance. This permits examination of the patient's heart during exercise and other usual physical and psychologically stressful situations in the patient's life. Magnetic tape recording is also used which permits later analysis of ECGs at high speed, or for spot checking. Digital computer programs also are available for immediate ECG analysis. Because of cost and complexity, this approach involves remote transmission from numerous ECG sources.

Impedance Cardiography. An electrical measure of the mechanical activity of the heart may be made by passing 4 to 5 milliamperes of 100-kHz current vertically through the heart, using electrodes at the neck and wrist. The change in this current as a function of the varying amount of blood in the heart through the cardiac cycle is measured with a pair of electrodes placed along the left and right midaxillary lines at the level of the heart (along the sides of the trunk about 4 inches below the armpits of an adult). Since the blood has appreciably lower resistivity than the body tissues, the impedance between the electrodes of approximately 25 ohms will be seen to decrease during diastole by about 0.1 ohm as blood flows into the heart and to increase sharply during systole as it is forced back out of the current path. Anomalies in the pumping cycle will show up if the blood moves out of the electrical field in an abnormal fashion.

Ballistocardiography. One important measure of heart activity is the force with which the heart ejects blood. If one stands quite still on a spring scale, one will note a small pulsation in the dial reading with each heartbeat. This ballistic recoil of the body as the blood is propelled from the heart can be measured more carefully if the subject lies on a specially constructed bed. Essentially, the bed becomes a platform which can be measured (motion) photoelectrically, or with a linear variable-differential transformer, or with a coil and magnet combination.

For analysis of cardiac output and mechanical heart function, the first- and second-derivative (velocity and acceleration) curves are generated as well as the curve of instantaneous position. These curves can be formed electronically by using high-pass filters, but such filters also increase the effect of noise in the system. To reduce this problem, some ballistocardiographs are made with a velocity pickup made by passing a magnet through a solenoid coil. The voltage appearing across such a coil will be proportional to the velocity with which the magnet is being moved through the coil. It then requires only one differentiation to obtain the acceleration curve and one integration to produce the motion curve, thus markedly improving the signal-to-noise ratio for the system.

Phonocardiography. Electronic stethoscopes have improved the efficiency of the strictly audio-type stethoscope. The latter instrument, of course, has been used for physicians for scores of years to detect heart sounds. An electronic system for heart-sound amplification should have a passband between 10 and 1,000 Hz. For the physician who has become accustomed to a conventional stethoscope, some retraining of the ear is required in switching to an electronic device.

Halter Monitoring. For a number of years, it has been possible to record the ECG of ambulatory patients on magnetic tape continuously over several hours for later playback and interpretation (halter monitoring). Only in recent years, however, has this method come into wide application. Such long-period ECGs are helpful in determining the etiology of dizzy spells or syncope of cardiac origin; studying the nature of palpitations and episodes of tachycardia; and assessing the effectiveness of anti-arrhythmic therapy; among other situations where a comparatively long time span of measurements is required to develop a complete picture.

See **Heart and Circulatory System (Human)** and the list of entries included at the end of that entry.

Additional Reading

Brandenburg, R. O., et al., Editors: "Cardiology: Fundamentals and Practice," Year Book, Chicago, Illinois, 1987.

Eagle, K. A., et al., Editors: "The Practice of Cardiology," Little Brown, Boston, Massachusetts, 1989.

Fisch, C.: "Electrocardiography of Arrhythmias," Lea and Feberger, Philadelphia, Pennsylvania, 1990.

Fowler, N. O.: "Diagnosis of Heart Disease," Springer-Verlag, New York, 1991.

Garson, A., Jr., Bricker, T., and D. G. McNamara: "The Science and Practice of Pediatric Cardiology," Lea and Febiger, Philadelphia, Pennsylvania, 1990.

Goldberger, A. L., and E. Goldberger: "Clinical Electrocardiography: A Simplified Apoproach," 4th Edition, Mosby-Year Book, St. Louis, Missouri, 1990.

Julian, D. G., et al., Editor: "Diseases of the Heart," Bailliere Tindall, London, 1989.

Levy, D., et al.: "Prognostic Implications of Echocardiographically Determined Left Ventricular Mass in the Framingham Heart Study," *New Eng. J. Med.*, 1561–1566 (May 31, 1990).

ELECTROCHEMICAL MACHINING (ECM). In this method of metalworking, electrical and chemical energy become the cutting edges of the tool. Electrical energy causes a chemical reaction which, in turn, dissolves metal from a workpiece into an electrolytic solution. The ECM tool (cathode) is brought very close to the workpiece (anode). The distance will range from less than 0.001 to 0.010 inch (0.025 to 0.25 millimeter). A low-voltage, high-density direct current passes between them through the electrically conductive electrolyte solution. This solution is pumped through the gap at pressures ranging up to 300 psi (20.4 atmospheres). The solution normally is maintained at about 100–120°F (38–49°C). The current that is passed through the electrolyte solution ranges widely. Machines using up to 20,000 amperes have been used. As the current passes from the workpiece to the tool, metallic particles on the surface of the workpiece (ions) are caused to go into solution. These particles are then swept away by the rapidly flowing electrolyte. Some of the advantages claimed for ECM include: (1) virtually no tool wear; (2) no burrs are produced; (3) both hard and soft metals can be machined at same rate; (4) usually additional finishing operations are not required; (5) no mechanical stresses are produced in workpiece surface; (6) no thermal effects are produced in the workpiece because of the relatively low temperature of the operation; (7) tough metals often can be removed faster by ECM than conventional methods; (8) good tolerances and repeatability are produced in complex as well as simple shapes; and (9) the process is easily automated. The basics of this process are also applied in deburring, grinding, and polishing operations.

ELECTROCHEMISTRY. That branch of science which deals with the interconversion of chemical and electrical energies, i.e., with chemical changes produced by electricity as in electrolysis or with the production of electricity by chemical action as in electric cells or batteries. The science of electrochemistry began about the turn of the eighteenth century.

Background. In 1796, Alessandro Volta observed that an electric current was produced if unlike metals separated by paper or hide moistened with water or a salt solution were brought into contact. Volta used the sensation of pain to detect the electric current. His observation was similar to that observed ten years earlier by Luigi Galvani who noted that a frog's leg could be made to twitch if copper and iron, attached respectively to a nerve and a muscle, were brought into contact.

In his original design Volta stacked couples of unlike metals one upon another in order to increase the intensity of the current. This arrangement became known as the "voltaic pile." He studied many metallic combinations and was able to arrange the metals in an "electromotive series" in which each metal was positive when connected to the one below it in the series. Volta's pile was the precursor of modern batteries.

In 1800, William Nicholson and Anthony Carlisle decomposed water into hydrogen and oxygen by an electric current supplied by a voltaic pile. Whereas Volta had produced electricity from chemical action these experimenters reversed the process and utilized electricity to produce chemical changes. In 1807, Sir Humphry Davy discovered two new elements, potassium and sodium, by the electrolysis of the respective solid hydroxides, utilizing a voltaic pile as the source of electric power. These electrolytic processes were the forerunners of the many industrial electrolytic processes used today to obtain aluminum, chlorine, hydrogen, or oxygen, for example, or in the electroplating of metals such as silver or chromium.

Since in the interconversion of electrical and chemical energies, electrical energy flows to or from the system in which chemical changes take place, it is essential that the system be, in large part, conducting or consist of electrical conductors. These are of two general types—electronic and electrolytic—though some materials exhibit both types of conduction. Metals are the most common electronic conductors. Typical electrolytic conductors are molten salts and solutions of acids, bases, and salts.

A current of electricity in an electronic conductor is due to a stream of electrons, particles of subatomic size, and the current causes no net transfer of matter. The flow is, therefore, in a direction contrary to what is conventionally known as the "direction of the current." In electrolytic conductors, the carriers are charged particles of atomic or molecular size called *ions*, and under a potential gradient, a transfer of matter occurs.

An electrolytic solution contains an equivalent quantity of positively and negatively charged ions whereby electroneutrality prevails. Under a potential gradient, the positive and negative ions move in opposite directions with their own characteristic velocities and each accordingly carries a different fraction of the total current through any one solution. Each fraction is referred to as the ionic transference number. Furthermore, the velocity increases with temperature causing a corresponding increase in electrolytic conductivity. This characteristic is opposite to that observed for most electronic conductors which show less conductivity as their temperature is increased.

The concept that charged particles are responsible for the transport of electric charges through electrolytic solutions was accepted early in the history of electrochemistry. The existence of ions was first postulated by Michael Faraday in 1834; he called negative ions "anions" and positive ones "cations." In 1853, Hittorf showed that ions move with different velocities and exist as separate entities and not momentarily as believed by Faraday. In 1887, Svante Arrhenius postulated that solute molecules dissociated spontaneously into *free ions* having no influence on each other. However, it is known that ions are subject to coulombic forces, and only at infinite dilution do ions behave ideally, i.e., independently of other ions in the solution. Ionization is influenced by the nature of the solvent and solute, the ion size, and solute-solvent interaction. The dielectric constant and viscosity of the solvent play dominant roles in conductivity. The higher the dielectric constant, the less are the electrostatic forces between ions and the greater is the conductivity. The higher the viscosity of the solvent, the greater are the frictional forces between ions and solvent molecules and the lower is the electrolytic conductivity.

In 1923, Debye and Hückel presented a theory which took into account the effect of coulombic forces between ions. They introduced the concept of the ion atmosphere, in which at some radial distance r from a central ion, there is, on a time average, an ionic cloud of opposite charge which sets up a potential field whose magnitude depends on the magnitude of r. This interionic attraction leads to two effects on the electrolytic conductivity. Under a potential gradient, an ion moves in a certain direction. However, the ion cloud, being of opposite sign will tend to move in the opposite direction, and because of its attraction for the central ion, will have a retarding effect on the ion velocity and thereby lead to a lowering in the electrolytic conductivity. On the other hand, the central ion will tend to pull the ion cloud with it to a new location. The ion atmosphere will adjust to its new location in time, but not instantaneously, and the delay results in a dissymmetry in the potential field around the ion. This also causes a lowering in the conductance of the solution. These effects become more pronounced as the concentration of the solution is increased; for dilute solutions, below about 0.1 molal, the equivalent conductance decreases with the square root of the concentration. For more concentrated solutions, the relation between conductivity and concentration is much more complex and depends more specifically on individual solute properties.

Interionic attraction in dilute solutions also leads to an effective ionic concentration or activity which is less than the stoichiometric value. The *activity* of an ion species is its thermodynamic concentration, i.e., the ion concentration corrected for the deviation from ideal behavior. For dilute solutions the activity of ions is less than one, for concentrated solutions it may be greater than one. It is the ionic activity that is used in expressing the variation of electrode potentials, and other electrochemical phenomena, with composition.

When electricity passes through a circuit consisting of both types of electrical conductors, a chemical reaction always occurs at their interface. These reactions are electrochemical. When electrons flow from the electrolytic conductor, oxidation occurs at the interface while reduction occurs if electrons flow in the opposite direction. These electronic-electrolytic interfaces are referred to as *electrodes*; those at which oxidation occurs are known as *anodes* and those at which reduction occurs, as *cathodes*. An anode is also defined as that electrode by which "conventional" current enters an electrolytic solution, a cathode as that electrode by which "conventional" current leaves. Positive ions, for example, ions of hydrogen and the metals, are called *cations* while negative ions, for example, acid radicals and ions of nonmetals are called *anions*.

Laws of Electrolysis. In 1833, Faraday enunciated two laws of electrolysis which give the relation between chemical changes and the product of the current and time, i.e., the total charge (coulombs) passed through a solution. These laws are: (1) the amount of chemical change. e.g., chemical decomposition, dissolution, deposition, oxidation, or reduction, produced by an electric current is directly proportional to the quantity of electricity passed through the solution; (2) the amounts of different substances decomposed, dissolved, deposited, oxidized, or reduced are proportional to their chemical equivalent weights. A chemical equivalent weight of an element or a radical is given by the atomic or molecular weight of the element or radical divided by its valence; the valence used depends on the electrochemical reaction involved. The electric charge on an ion is equal to the electronic charge or some integral multiple of it. Accordingly, a univalent negative ion has a charge equal in magnitude and of the same sign as a single electron, and its chemical equivalent weight is equal to its atomic weight, if an element, or to its molecular weight, if a radical. A trivalent ion has $+3$ or -3 electronic charges, depending on whether it is a positive or a negative trivalent ion. For trivalent ions, then, the equivalent weight would be equal to its atomic weight, if an element, or to its molecular weight, if a radical, divided by three.

The quantity of electricity required to produce a gram-equivalent weight of chemical change is known as the *faraday*. A faraday corresponds, then, to an *Avogadro number of charges*. The most accurate determination of the faraday has been made by a silver-perchloric acid coulometer in which the amount of silver electrolytically dissolved in an aqueous solution of perchloric acid is measured. This method gives 96,487 coulombs (or ampere-seconds) per gram-equivalent for the faraday on the unified ^{12}C scale of atomic weights adopted in 1961 by the International Commission on Atomic Weights.

Electrochemical Equivalent. Preferably termed *coulomb equivalent* of an element or radical, this is the weight in grams which is equivalent to 1 coulomb of electricity and is given by the gram-equivalent weight divided by the faraday (96,487 coulombs per gram-equivalent); for example, the electrochemical equivalent of silver is given by 107.870/96,487 or 0.00111797 grams/coulomb where 107.870 is the atomic weight of silver based on the unified ^{12}C scale adopted in 1961. The electrochemical equivalents of other elements may be calculated in like fashion.

In electrolysis and in any electric cell or battery, there is an electromotive force (emf) or voltage across the terminals. This emf is expressed in the practical unit, the volt, which is equal to the electromagnetic unit in the meter-kilogram-second system. In any one cell, the emf is the sum of the potentials of the two electrodes and of any liquid-junction potentials that may be present. Neither of the individual electrode potentials can be evaluated without reference to a chosen reference electrode of assigned value. For this purpose, the hydrogen electrode has been universally adopted and is arbitrarily assigned a zero potential for all temperatures when the hydrogen ion is at unit activity and the hydrogen gas is at atmospheric pressure. A hydrogen electrode consists of a stream of hydrogen gas bubbling over platinized platinum or gold foil and immersed in a solution containing hydrogen ions; the electrochemical reaction is: $\frac{1}{2}h_2(\text{gas}) = H^+$ (solution) $+ \epsilon$, where ϵ represents the electron. The potential of the hydrogen electrode, E_H, as a function of hydrogen ion concentration and hydrogen-gas pressure is given by

$$E_H = E_H^0 - (RT/nF)\ln(a_{H^+}/p_{H2}^{1/2})$$

$$= E_H^0 - (RT/nF)\ln(c_{H^+}f_{H^+}/p_{H2}^{1/2}),$$

where E_H^0 is the standard quantity assigned a value of zero, R is the gas constant, T the absolute temperature, n the number of equivalents, F the faraday, p_{H2} the pressure of hydrogen, and a_{H^+}, c_{H^+} and f_{H^+}, respectively, the activity, concentration, and activity coefficient of hydrogen ions. When a_{H+} and $p_{H2}^{1/2}$ equal one, $E_H = E_H^0$. For very dilute solutions below 0.01 molal f_{H+} may be taken as unity without appreciable error.

The standard potentials, E^O, of other electrodes are obtained by direct or indirect comparison with the hydrogen electrode. Values are determined at 25°C. The values for several metals and other elements are given in entry on **Activity Series**. The reducing power of the elements decreased on going down the column from those elements with negative standard electrode potentials to those with positive potentials. These values are for the ions at unit activity, and reversible or thermodynamic values as a function of metal or radical concentration are given by equations similar to the one above. For the general reaction: $M = M^{n+} + n\epsilon$, the potential is given by $E_m = E_H^0 - (RT/nF)1n\, a_{Mn+}$.

In electrolysis, at very low current densities, the potentials of the electrodes approximate in magnitude their reversible values and deviate somewhat from these values because of an *IR* drop in the solution and possible concentration polarization (the concentration at the electrode surface may differ from that in the bulk of the solution). Also for high current densities, especially for the generation of gases such as hydrogen, oxygen or chlorine, the voltage required exceeds the reversible voltage; the excess voltage is known as overvoltage, or overpotential for a single electrode, and arises from energy barriers at the electrode. Overpotential, in general, increases logarithmically with an increase in current density.

Scope of Electrochemistry. In addition to what has previously been described, it is customary to include under electrochemistry: (1) processes for which the net reaction is physical transfer, e.g., concentration cells; (2) electrokinetic phenomena, e.g., electrophoresis, electro-osmosis, streaming potential; (3) properties of electrolytic solutions if determined by electrochemical or other means, e.g., activity coefficients and hydrogen ion concentration; (4) processes in which electrical energy is first converted to heat which in turn causes a chemical reaction to occur that would not do so spontaneously at ordinary temperature. The first three are frequently considered a portion of physical chemistry, and the last one is a part of electrothermics or electrometallurgy.

The passage of electricity through gases is sometimes included under electrochemistry. However, in electrical discharges in gases, the principles are entirely different from what they are in the electrolysis of electrolytic solutions. Whereas in the latter, ionic dissociation occurs spontaneously as a result of forces between solvent and solute and without the application of an external field, for gases relatively high voltages must be applied to accelerate the electrons from the electrode to a velocity at which they can ionize the gas molecules they strike. In this case, the resulting chemical reaction taking place between ions, free radicals, and molecules occurs in the gas phase and not at the electrodes as in the electrolysis of solutions. Studies of the electrical conduction of gases, accordingly, are generally considered under the physics of gases.

Electrochemistry finds wide application. In addition to industrial electrolytic processes, electroplating, and the manufacture and use of batteries already mentioned, the principles of electrochemistry are used in chemical analysis, e.g., polarography, and electrometric or conductometric titrations; in chemical synthesis, e.g., dyestuffs, fertilizers, plastics, insecticides; in biology and medicine, e.g., electrophoretic separation of proteins, membrane potentials; in metallurgy, e.g., corrosion prevention, electrorefining; and in electricity, e.g., electrolytic rectifiers, electrolytic capacitors.

Additional Reading

Adams, R. N.: "Electrochemistry at Solid Electrodes," McGraw-Hill, 1969. A *classic reference*.

Alkire, R. and D-T. Chin, Editors.: "Tutorial Lectures in Electrochemical Engineering and Technology (II)," American Institute of Chemical Engineers, New York, 1983.

Bard, A. J., et al.: "Standard Potentials in Aqueous Solution," McGraw-Hill, New York, 1985.

Kissinger, P. T., and W. K. Heineman: "Laboratory Techniques in Electroanalytical Chemistry," McGraw-Hill, New York, 1984.

Law, C. G., Jr. and R. Pollard, Editors: "Processing of Electronic Materials," American Institute of Chemical Engineers, New York, 1987.

Lund, H., and M. M. Baizer: "Organic Electrochemistry: An Introduction and Guide," 3rd Edition, McGraw-Hill, New York, 1991.

Mattson, J. S., et al.: "Electrochemistry: Calculations, Simulation, and Instrumentation," McGraw-Hill, New York, 1972.

Trescott, M. M.: "The Rise of the American Electrochemicals Industry, 1880-1910," Greenwood, Westport, Connecticut, 1981.

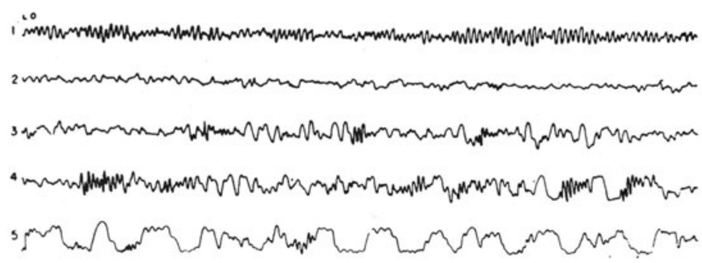
Vijh, A.: "Electrochemistry of Metals and Semiconductors: The Application of Solid State Science to Electrochemical Phenomena," McGraw-Hill, New York, 1973.

White, R. A., Savinell, R. F., and A. Schneider: "Electrochemical Engineering Applications," American Institute of Chemical Engineers, New York, 1987.

ELECTRODE. In an electric circuit, part of which is composed of other than the usual conductor of copper, or other metal, the terminal connecting the conventional conductor and the conducting substances is an electrode. Examples of electrodes are to be found in the electric cell, where they dip in the electrolyte; the electric furnace, where the electrodes connect the external circuit with the heating arc; and the metallic elements in thermionic tubes and gas-discharge devices, and in semiconductor devices, where they perform one or more of the functions of emitting, collecting or controlling by an electric field the movements of electrons and ions. See also **Graphite.**

ELECTRODIALYSIS. The removal of electrolytes from a colloidal solution by a combination of electrolysis and dialysis. Usually the colloidal solution is placed in a vessel with two dialyzing membranes with pure water in compartments on the other side of the membranes. Two electrodes are inserted in the pure water compartments and an applied emf causes the ions to migrate from the colloidal solution. See also **Desalination;** and **Ocean Resources (Energy).**

ELECTROENCEPHALOGRAM. A graphic tracing of the electric impulses of the brain, sometimes abbreviated EEG. The instrument which makes the record is known as an electroencephalograph. The electrical activity of the brain is manifested at the surface of the scalp by small potential changes on the order of 5 to 200 microvolts in a frequency band from 1 to 50 Hz. A sample of so-called brain waves or electroencephalogram is shown by the accompanying diagram. When viewed casually, the usual EEG would appear to be simply random noise, but spectral analysis and some direct electroencephalograms show pronounced components at several frequencies. The lowest (1 to 3 Hz) is termed the *delta rhythm*. Next highest are the *theta waves* (4 to 7 Hz). The *alpha rhythm* is the most pronounced and occurs between 8 and 13 Hz. A third pronounced rhythm appears between 13 and 30 Hz and is termed the *beta rhythm*. As shown by the diagram, the alpha rhythm is most pronounced during light sleep, while the delta rhythm appears during deep sleep. One of the uses of the EEG is determination of stages of sleep in sleep research and associated investigations.



EEG tracings of normal subject during sleep. Numbers shown at left indicate increasing depth of sleep.

In several diseases of the brain, especially epilepsy and cerebral tumors, very slow delta waves of higher-than-average voltage may be seen. Occasionally, as in epileptic seizures, similar waves of very high frequency may occur. In tumors, the appearance of such waves is confined to the area immediately surrounding the tumor; hence electroencephalography has some ancillary value in the localization of cerebral tumors. See **Seizure (Neurological).**

Electroencephalographic techniques also have been used rather extensively in research on sleep and sleep apnea associated with insomnia.

A conventional electroencephalogram is taken by placing a series of 6 to 10 electrodes symmetrically on each side of the head from the front to the back, with a reference electrode on the mastoid bone or some other similar spot. Sometimes the electrodes consist of small pins which puncture the skin to make better contact, but usually the scalp electrode is a small sponge soaked with saline solution and held down with tape. Usually, it is not necessary to shave the hair from the scalp. The usual EEG machine uses a multichannel, direct-writing recorder with 6 to 10 channels and speeds ranging from 2 to 50 millimeters per second, the width per channel typically being about 25 millimeters. Because the EEG amplitude is an order of magnitude smaller than the electrocardiograph amplitude, the criteria for the design of an EEG amplifier apply much more strenuously. See **Electrocardiography**. Common-mode rejection should be at least 10,000:1, and the recording should be done in a shielded room. In contrast, low-frequency response need not be as good for the EEG as for the ECG; thus polarization of the electrodes is not so critical. When making an EEG recording, it is necessary that the shielded room be relatively free of external stimuli, such as changes in ambient light, sound, or vibration because sudden changes may produce startle-response artifacts.

A technique known as the evoked-cortical-potential makes it possible to see the effect on the EEG of a sensory stimulus. At the same time as the stimulus is initiated (such as a flash of light or burst of sound), an EEG recording is made and stored in a computer memory, usually on a digital basis. Each time a fresh stimulus is applied, the storage is updated, so that after 50 or 100 trials, the digital memory contains an average EEG synchronized with the stimulus. A typical computer memory for this application may contain 2,500 bits; 50 elements in one direction could represent fifty 10-millisecond periods, starting with the onset of the stimulus, while the 50 elements in the orthogonal direction would represent the average amplitude of the EEG for each time period. This technique is useful for research work on sight, sound, touch, and other sensory stimuli and, under some circumstances, in the location of disorders causing blindness or deafness.

Work on the brain and on individual fibers frequently requires the use of microelectrodes to stimulate and detect the responses from very tiny spots. Microelectrodes, consisting of very fine wires surrounded by thin glass insulation, are made down to an active diameter of approximately one micrometer. The impedance of such an electrode can be 10 to 100 million ohms, and the frequency of the signal can have components up to 10 kHz. At the same time, the signal-to-noise ratio can be very low. Thus, the requirements of very high common-mode impedance, very high input impedance, and virtually no input capacitance dictates a special class of amplifier in which the input capacitance is neutralized by feedback. Apparatus is available that allows the simultaneous use of an electrode for producing a stimulus and recording the response. The input and output circuits to the electrode must be carefully isolated to make this possible.

ELECTROENDOSMOSIS. Electrophoresis in which the solid is stationary and the water phase is displaced and migrates toward the electrode.

ELECTROFORMING. The electrolytic deposition of metal upon a conducting mold, to make a desired metal object, such as precision tubing or medals. The mold is often of graphite-coated wax, so that it can be removed by melting. See **Electroplating.**

ELECTROKINETIC EFFECTS. Movements of particles under the influence of an applied electric field.

ELECTROKINETIC TRANSDUCER. A transducer that depends for its operation on the dielectric polarization in certain liquids resulting from viscous shearing stress that accompanies flow through porous materials.

ELECTROKINETIC (ZETA) POTENTIAL. The difference in potential between the immovable liquid layer attached to the surface of a solid phase and the movable part of the diffuse layer in the body of the liquid.

ELECTROLUMINESCENCE. See **Luminescence.**

ELECTROLYSIS. See **Electrochemistry.**

ELECTROLYSIS-TYPE CHEMICAL ANALYZER. Sometimes referred to as electroplating analyzers, these devices can be used for determining metals and other materials that will plate out on an electrode which is part of an electrolytic cell. Alloys, such as stainless steel, brass, and bronze that contain chromium, copper, lead, iron, nickel, and tin-bearing metals containing bismuth, can be analyzed in this fashion. The sample must be relatively easy to dissolve so that an electrolyte can be formed and it must be sufficiently large to permit plating out a quantity of material that can be accurately weighted. The potential required to effect plating of a specific material should be known in advance. Complex materials usually are identified on a cumulative basis by making stepwise increases in plating potential, with intermittent weighing of the plated electrode.

ELECTROLYTE. See **Battery, Ionic Mobility; Kohlrausch Law.**

ELECTROLYTIC CONDUCTIVITY AND RESISTIVITY MEASUREMENTS. Industrial interest in the measurement of electrolytic conductivity (of which electrolytic resistivity is the reciprocal) arises chiefly from its usefulness as a measure of ion concentrations in water solutions. Also, by comparison with other analytical methods, this is relatively simple and inexpensive.

Pure water is a very poor conductor. Water, such as may be produced by passage through a mixed-bed ion exchanger, has a conductivity approaching very closely the theoretical minimum of approximately 0.05 μS/cm (18 MΩ·cm) at 25°C, which is due to the dissociation products of water itself (S/cm = siemens/centimeter). The conductivity of a water solution, as encountered in industrial practice, is almost exclusively due to a dissolved electrolyte rather than to the solvent (water) ions, and thus a criterion for electrolyte concentration can be established. Solutions of strong electrolytes follow a rather uniform pattern of change in conductivity with concentration, which is almost linear at low concentrations, rising more gradually to a maximum (usually about 20-30% wt) and then falling as the concentration rises further. A series of such conductivity-concentration curves is given in Fig. 1.

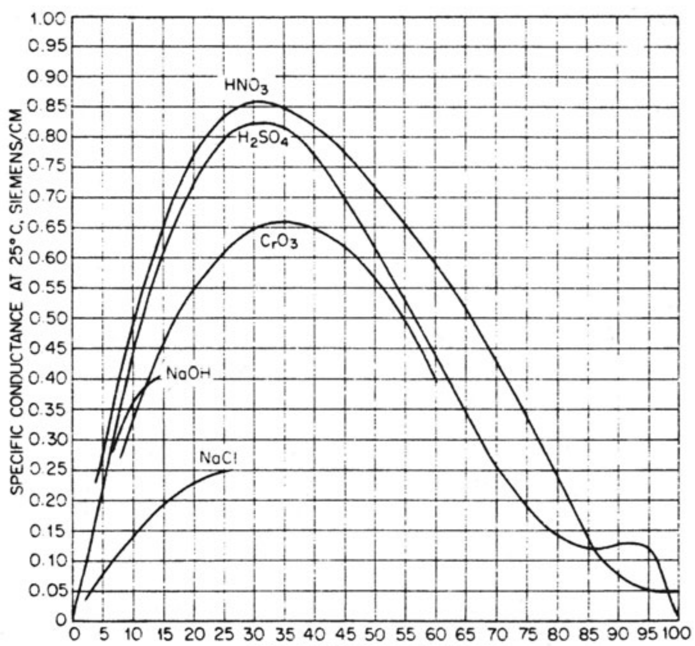


Fig. 1. Conductivity-concentration curves for certain electrolytes.

Practical applications of these measurements include: (1) Gauging the quality of pure water, such as distilled or demineralized water and condensed boiler steam; (2) measuring the extent of reactions, such as neutralization, precipitation, and washing of soluble electrolytes from insoluble materials; (3) detecting contamination, such as leaks in heat exchangers and resultant contamination of heating or cooling media as in acid coolers, condenser coils, and steam coils; (4) checking possible saltwater intrusions in streams and wells; (5) checking on process interface levels where, for example, oil-in-water and water-in-oil emulsions may be distinguished by the fact that the former are conductive and the latter are essentially nonconductive; and (6) the enhancement of the conductivity of certain ions of interest. For example, very conductive hydroxyl ions may be removed from boiler water by the addition of a weakly ionized organic acid to reveal a conductivity which is more nearly proportional to that of the remaining dissolved salts. Or, conductive ammonium hydroxide may be removed from steam condensate by passage through a hydrogen cation exchanger to reveal the conductivity of the remaining dissolved salts, which are converted to their corresponding mineral acids. Similarly, samples may be boiled or sparged to remove conductive dissolved gases.

Measurement Fundamentals

The flow of electricity through matter is accomplished by movement of electric charges, which in metallic conductors are electrons and in electrolytic conductors are positive and negative ions. Conducting solutions in general are electrolytic conductors. In electrolytic conductors, current is usually introduced and leaves the system through metallic electrodes on the surface of which chemical reactions occur. It is possible when using alternating current to cause current to flow by inductive coupling as well as by direct contact between electrode and electrolyte. Positive ions or cations move toward the cathode, where reduction takes place, and negative ions or anions move toward the anode, where oxidation takes place. The conductivity of a solution depends upon the concentration and mobility of all ions present. The ion mobility, in turn, depends upon ion size and charge, as well as the dielectric constant of the solvent and the solution temperature and viscosity.

The determination of electrolytic conductivity presently consists of measuring the ac electrical conductance of a column of solution. Although ac measurement methods greatly reduce errors associated with electrolysis, when electrodes are used, they introduce other errors associated with series and shunting capacitance which must be compensated for in the design of the measuring instrument. A precision of 0.01% can be obtained in laboratory measurements following the bridge techniques first discussed by Grinnell Jones and his coworkers. Industrial conductivity meters are capable of providing accuracies of 1% of the actual conductivity under ideal conditions.

Electrolytic conductivity is most often measured by placing electrodes in contact with the electrolytic solution which is contained in such a way that the measured electrical conductance between the electrodes can be related to the conductivity of the solution. The conductivity cell most commonly comprises an enclosure made of electrically insulating material, such as glass or plastic, which serves to hold or isolate a portion of the electrolytic solution and to accommodate the two electrodes. The cell constant of such a device is then used to relate the measured electrical conductance between the electrodes to the actual electrolytic conductivity.

Two electrodes, 1 centimeter square, located on opposite interior faces of a hollow cube, 1 centimeter on an edge, would have a cell constant of 1/cm, and a measured conductance of 100 microsiemens at 25°C would indicate a conductivity of 100 microsiemens/cm (10 millisiemens/m) at 25°C.

Definitions and Units. Electrolytic conductivity is often defined as the electrical conductance of a unit cube of solution as measured between opposite faces. It is expressed in the same units as electrical conductivity, i.e., reciprocal ohms per unit length. Most commonly we find Mho/centimeter (Ω^{-1} cm^{-1}), siemens/centimeter (S cm^{-1}), and siemens/meter (S m^{-1}):

$$1\ \Omega^{-1}\ \text{cm}^{-1} = \text{S cm}^{-1} = 100\ \text{S m}^{-1}.$$

Few solutions exhibit conductivities as great as 1 siemens/cm. The most commonly used decimal submultiples are micromho/centimeter ($\mu\Omega^{-1}$ cm^{-1}), microsiemens/centimeter (μS cm^{-1}), and millisiemens/meter (mS m^{-1}):

$$1\ \mu\Omega^{-1}\ \text{cm}^{-1} = 1\ \mu\text{S cm}^{-1} = 0.1\ \text{mS m}^{-1}.$$

Electrolytic resistivity (the reciprocal of conductivity) is similarly defined as the electrical resistance of a unit cube of solution. It is expressed in the same units as electrical resistivity, i.e., ohms times a unit of length. Most commonly we find: ohm-cm (Ω-cm) and ohm-meter (Ω-m):

$$100\ \Omega\text{-cm} = 1\ \Omega\text{-m}.$$

Again, decimal multiples commonly encountered are megohm-centimeter and megohm-meter:

$$100\ \text{M}\Omega\text{-cm} = 1\ \text{M}\Omega\text{-m}.$$

Resistivity units are used almost exclusively to describe ultrapure water in the 10 megohm-cm to 18 megohm-cm (0.1 microsiemens/cm to 0.055 microsiemens/cm) range generated by mixed bed ion exchange and used as boiler feed water and in certain critical washing applications.

The cell constant of a conductivity cell is defined as a factor which relates the measured conductance between the cell terminals to the conductivity of the electrolyte being measured. It is generally expressed in reciprocal units of length (although occasionally in units of length for certain European manufacturers). Most commonly we find 1/centimeters (cm^{-1}) and 1/meters (m^{-1}):

$$1\ \text{cm}^{-1} = 100\ \text{m}^{-1}.$$

The conductance measured between the cell terminals is multiplied by the cell constant given in reciprocal units of length to calculate the conductivity. The measured resistance between the cell terminals is divided by the cell constant to calculate the resistivity. Although the cell constant in reciprocal units of length can be calculated from the dimensions of the conductivity cell by dividing the length of the electrical path through the solution by the cross-sectional area of the path, in practice, these measurements are difficult to make and are only used to approximate the cell constant, which is determined by use of standard solutions of known conductivity or by comparison with other conductivity cells which have been so standardized.

Measuring Circuits

Although there are several circuits used for measuring electrolytic conductivity, the ac Wheatstone bridge is widely applied and is potentially the most stable and accurate.

AC Wheatstone Bridge. A typical system is shown in Fig. 2 and comprises the bridge, including the voltage source, the null indicator, and the conductivity cell. In Fig. 2, *D* represents an ac voltage-sensitive device called the *detector*. The ac source may be the low-voltage tap on a line-frequency operated transformer or battery or line-powered elec-

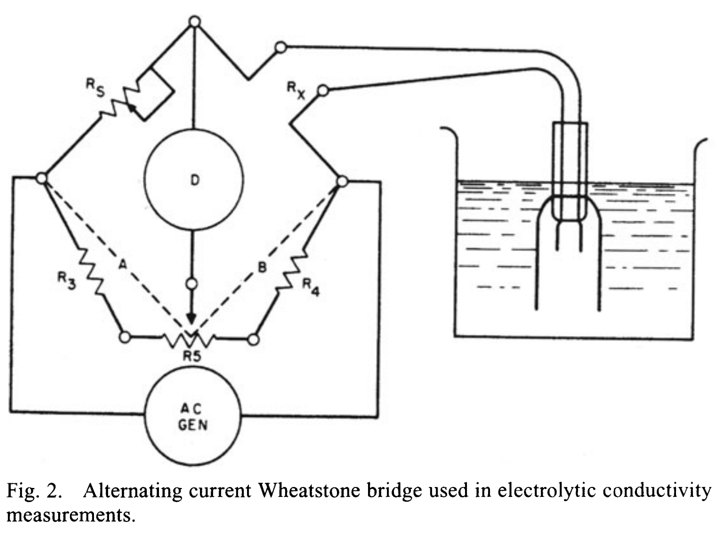


Fig. 2. Alternating current Wheatstone bridge used in electrolytic conductivity measurements.

tronic oscillator for higher frequencies. The magnitude of the bridge voltage necessarily is related to the sensitivity of the detector and also to the general characteristics of the electrolytes to be tested.

The usual industrial measuring and control equipment is supplied with bridge voltages of 1–10 V. The frequency of this ac source in commercial units is commonly 60 Hz and more rarely 1000 Hz. Where measurements are to be made on high-resistance electrolytes, such as distilled water or steam condensate, the lower bridge source frequency is preferable. For measurements in high-conductivity solutions, the higher bridge frequencies are of advantage.

R_S is the so-called *standard arm* of the bridge and is generally made variable, as a device either to change the range of the instrument by selecting one of a number of resistors differing in resistance by powers of 10 or to correct for the temperature coefficient of resistance of the electrolyte. R_3 and R_4 are end resistors whose function is to establish the limits of the bridge calibration. R_5 is the calibrated slidewire potentiometer. With R_3 and R_4 short-circuited, the range of the bridge would be zero to infinity in resistance or conductance. Increasing values of R_3 and R_4 compared with the value of R_5 will reduce the range covered. It should be noted that the slidewire contact resistance is in series with the detector, and thus variable values of this resistance cause no error in bridge readings. R_X is effectively the resistance of the electrolyte measured between the two electrodes of the conductivity cell immersed in the liquid under test. The condition for balance of the Wheatstone bridge is that $A/B = R_S/R_X$, and this condition is indicated by no current flow through the detector D.

While most laboratory conductivity bridges are manually balanced, the Wheatstone bridge circuit also finds use in a variety of conductivity monitors, controllers, and recorders where it is mechanically rebalanced by a servomechanism operated by the detector. Generally in these devices, advantage is taken of the phase shift which occurs in the detected signal as the bridge is driven through balance by the servo motor.

Conductivity Meter. A second system utilizes a simple ohmmeter circuit, shown in Fig. 3. A meter, transformer secondary winding, and conductivity cell are connected in series so that the current is a function of the cell conductance. The meter may be calibrated in resistivity or conductivity units.

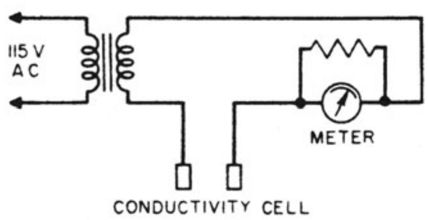


Fig. 3. Conductivity measurement system utilizing a simple ohmmeter circuit.

While early circuits of this type suffered from inaccuracies due to line voltage variations, the addition of a regulated power supply to drive the transformer has brought this relatively simple and inexpensive circuit into wide use. Complete isolation may be achieved by interposing a second transformer between the cell and meter. Generally, a stage of amplification is added to increase sensitivity and to reduce nonlinearity caused by meter resistance. This, combined with gated detection, reduces those polarization errors associated with series capacitance at the electrodes. Driven shields are employed to reduce the errors associated with the shunt capacitance of long cell leads. The addition of automatic temperature compensation, alarm contacts, and electrical outputs make the conductivity meter the most widely used instrument for industrial measurement and control applications.

Electrodeless Circuit. An electric current may be caused to flow in an electrolyte by means of induction without the use of contacting electrodes. In such electrodeless systems, the electrolyte is contained in an electrically insulating tube which passes through the cores of two transformers (Fig. 4) in such a way that the electrolyte forms a closed loop linking the flux in both cores. In the first transformer, this loop of electrolyte serves as a single-turn secondary winding in which an alternating voltage is induced. In the second transformer, the loop forms a single-turn primary winding, providing a means for measuring the resulting current which is directly proportional to the specific conductance of the electrolyte comprising the loop. Alternatively, both transformers may be located about an insulated tube immersed in the electrolytic solution.

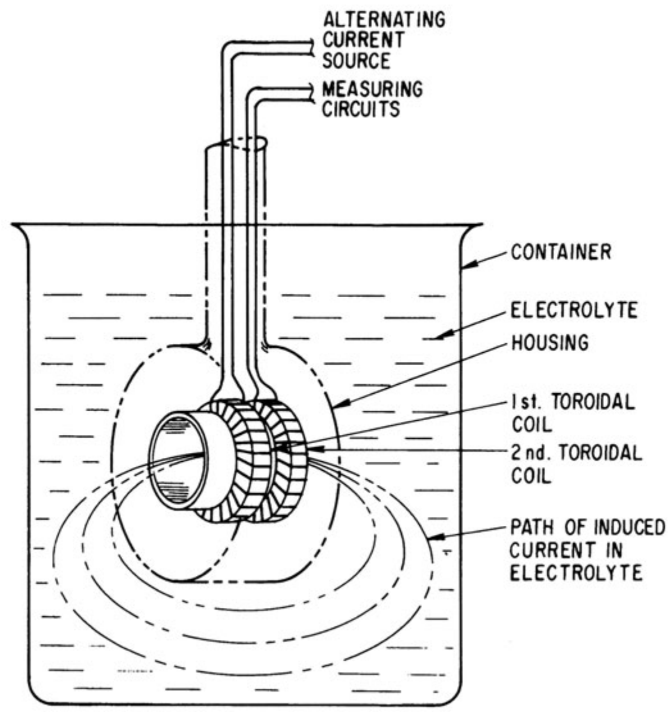


Fig. 4. Inductive electrolytic conductivity measuring circuit.

Variations of these systems employing glass tubes were in use before 1907. However, more recently, the introduction of chemically resistant, high-temperature electrical insulators, such as the fluorocarbons, has simplified the design of the insulated tube comprising the electrodeless conductivity cells. Since no contacting electrodes are used, all the problems associated with electrodes, such as polarization and electrode surface maintenance, simply disappear. Wide application of these systems is found in highly conductive electrolytes, such as the strong mineral acids and bases—often in conjunction with abrasive slurries or materials containing entangling fibers.

Four-Electrode Circuit. This method avoids errors caused by polarization and fouling by using a set of measuring electrodes located between a set of current-producing electrodes. See Fig. 5. The measuring electrodes are used to determine the voltage drop in the electro-

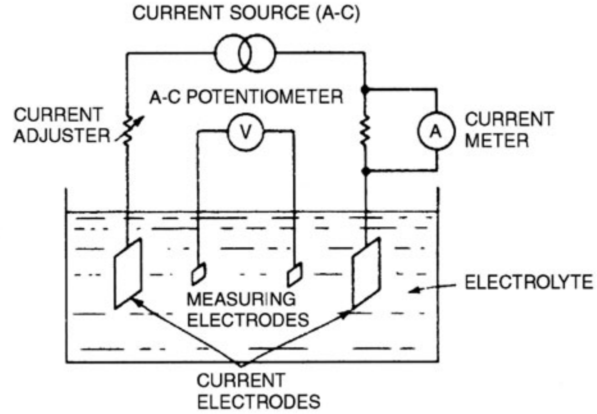


Fig. 5. Four-electrode conductivity circuit.

lyte caused by the current. The conductivity of the electrolyte between the measuring electrodes is proportional to the current divided by the potential. Laboratory measurements may be made with either ac or dc. However, process instrumentation almost always utilizes ac. In practice, the alternating current through the entire cell is varied to maintain a constant potential between the measuring electrodes. When this is done, the conductivity of the electrolyte between the measuring electrodes is proportional to the cell current. Changes external to the measuring electrodes, such as may be caused by current electrode polarization or fouling, will not cause a change in the cell current, which will be maintained at such a value that the potential between the measuring electrodes is constant. Systems have been designed that can accommodate a tenfold increase in impedance at the current electrodes due to polarization and fouling. Fouling and polarization errors do not occur at the measuring electrodes since the measurement there is essentially potentiometric with no current flowing through the measuring circuit. Size and orientation of the measuring electrodes are also chosen to avoid such errors.

Conductivity Cells

These are simple in basic structure, consisting typically of two metal plates or electrodes spaced within an insulating chamber. Examples are shown in Figs. 6 and 7. This arrangement permits isolation and measurement of a portion of the solution and serves to make the measured resistance independent of sample volume and proximity to conductive and nonconductive surfaces. In laboratory cells, platinum electrodes mounted in a glass are commonly employed for their excellent chemical resistance.

Dip Cell. This is designed for dipping or immersing into open vessels. See Fig. 6.

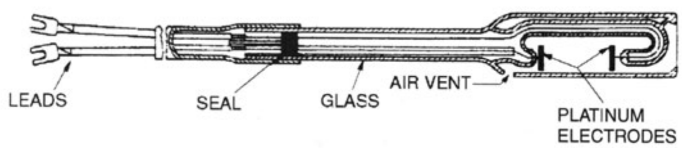


Fig. 6. Dip-type conductivity cell.

Screw-in Cell. This is designed for permanent installation in pipelines and tanks and is equipped with threaded fittings. See Fig. 7.

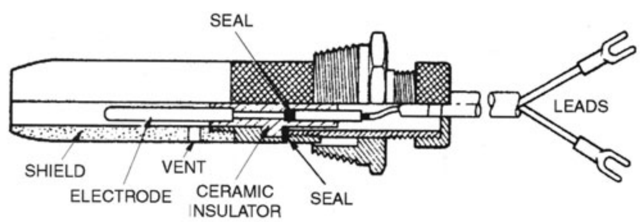


Fig. 7. Screw-in conductivity cell for high-pressure service.

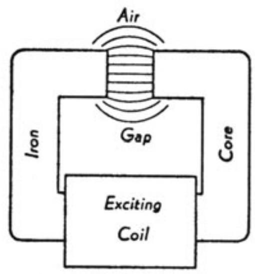
Insertion Cell with Removal Device. This is configured to permit removal of the element without closing down or depressurizing the line in which it is installed.

Flow Cell. This type of cell is built in sections of plastic or glass tubing with bores from several millimeters to one inch (25.4 mm) and more. The cell has internal electrodes, usually metallic or carbon rings, mounted flush with the wall to offer little resistance to flow.

Elmer Sperry and John Nagy, Beckman Industrial Corporation.

ELECTROLYTIC TRANSDUCER. A device whose electrolytic resistance is caused to change when exposed to the variable being measured. A simple configuration comprises two electrodes that are separated by a small distance, with the volume between the electrodes occupied by an electrolyte. To be used as a transducer, as in the case of displacement measurement, the distance between the electrodes or the effective cross-sectional area of the electrolyte can be varied. Very little force is required to produce a displacement and the device can be made quite small. The devices are very temperature sensitive, a disadvantage for practical, industrial applications. Also, like other electrolytic cells, the devices are subject to polarization. Because of these limitations, electrolytic transducers are not commonly used.

ELECTROMAGNET. A magnet whose field is produced by an electric current, and which is largely demagnetized upon cessation of the current, is an electromagnet. In order to obtain the strongest field possible, highly permeable soft iron or steel is employed for the core of electromagnets. In an electromagnet the current flows through a solenoid, which is a conductor wound in the form of a helix, and which produces a strong magnetic field coaxial with the helix. The core is placed inside the helix in order to give a magnetic path of the least reluctance. Electromagnets are found in a number of different forms, such as the plain solenoid with cylindrical core, or the horseshoe electromagnet, much used in electric bells, telegraph instruments, and telephones. Very powerful electromagnets are often used to move masses of iron, such as scrap iron, and have the advantage that the loading or unloading of the crane to which the magnet is attached is simply a matter of applying or disconnecting the electric current.



Electromagnet having two poles, one of a large variety of available designs.

The use of superconductors can significantly increase the effectiveness of an electromagnet.

ELECTROMAGNETIC PHENOMENA. The term *electromagnetic* is used to describe the combined electric and magnetic fields that are associated with movements of electrons through conductors; to the combined electrical and magnetic effects exhibited by and used by equipment, apparatus, and instruments; and, in terms of radiation, to describe the radiation that is associated with a periodically varying electric and magnetic field that is traveling at the speed of light, such as light waves, radio waves, x-rays, gamma radiation, and so on.

Electromagnetism. The pioneer discovery of the magnetic effect of the electric current was made by Oersted at Copenhagen in 1820. In experimenting with battery currents, he happened to bring a compass needle near a wire in which there was an electric current, and noted that the needle was deflected. Such a wire is surrounded by a magnetic field so that, to one looking along the wire in the direction from the positive to the negative battery terminal (the so-called "direction of the current"), the direction of the field, as indicated by the north pole of the compass needle, is clockwise (Ampère's rule).

If the wire carrying the current is placed in a magnetic field perpendicular to its direction, this field reacts with that due to the current in such a way as to give the wire a lateral thrust, perpendicular to both the wire and the field in which it is placed. For a wire of length l carrying current I and placed across a field of intensity B, this lateral force is given by the equation $f = BIl$, which follows from the definition of the ampere. An electric motor is driven by forces thus produced.

If the wire is bent into a circular loop of radius r, still carrying current I, there is produced at its center, perpendicular to the plane of the loop, a magnetic field of intensity $H = I/2r$. This, and the statement in the preceding paragraph, may be shown to be interdependent. If more loops are added, forming a coil of n equal turns close together, the resulting field is n times as great. By winding the n turns

along a cylinder, forming a "helix" of radius r and axial length a, one obtains something greatly resembling a bar magnet, the ends of the helix corresponding to the poles. The field intensity at the center of the axis of this helix is
The above expressions are all appropriate in the rationalized mksa system.

$$H_0 = \frac{nI}{\sqrt{4r^2 + a^2}}$$

If we now insert an iron core, we have an electromagnet, and the helix supplies the magnetomotive force nI ampere-turns for a magnetic circuit composed partly of iron and partly of air.

More general calculations of electromagnetic effects are based upon Ampère's law, the Biot-Savart law, and Maxwell's equations.

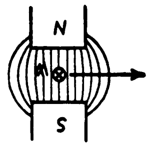
Electromagnetic Induction. Probably the most noteworthy of the many scientific contributions of the renowned Michael Faraday was his discovery in 1831 of electromagnetic (or, more logically, magnetoelectric) induction. As exhibited in the usual experimental arrangements, this phenomenon is the setting up, in a circuit, of an electromotive force by reason of the variation of the magnetic flux linked with the circuit; the magnitude of that electromotive force being, as Faraday found, proportional to the rate at which the flux through the circuit, or the "linkage," varies. If the flux linkage with the circuit, in weber-turns, is expressed by $N\phi$ (the actual flux, ϕ, times the number of turns, N), the electromotive force generated by its variation, in volts, is

$$E = N\frac{d\phi}{dt}$$

The electromotive force is positive (counterclockwise) when $d\phi/dt$ is positive, that is, when the flux is increasing, negative when it is decreasing; as viewed by one looking in the direction of the magnetic induction.

Another aspect of the matter is that if a conductor moves through a magnetic field, or if a magnetic field sweeps over a conductor, in such a way that the conductor cuts across the lines of force, the electricity in the conductor experiences forces at right angles to the field and to the (relative) motion. More general still is the Maxwell concept that when magnetic lines of force move sidewise, their movement results in an electric field at right angles to the magnetic lines and to their motion.

Faraday's discovery was almost accidental. Happening to thrust a bar magnet into a coil connected with a galvanometer, he noted a momentary deflection of the needle. If the north pole is thrust downward into the coil, so as to increase the flux linked with the coil, the current will be counterclockwise as viewed from above, and reverses on drawing the magnet out again. See accompanying figure.



Wire (W) moving to the left across an upward magnetic field has induced in it an emf away from the observer.

The far-reaching consequences of this simple observation can hardly be overestimated. It was the forerunner of the invention of electric generators and alternators, of the original Bell telephone, of the induction coil, of the transformer, of the induction motor, of magnetic damping devices, and of many other electric appliances. It is the basis of Lenz's law and of the Wilson experiment, and the explanation of eddy currents. The volt and the henry are definable in terms of it. The phenomenon is called mutual induction when the variation of current in one circuit causes a variation of magnetic flux through, and hence an electromotive force in, another coupled circuit. It is a curious fact that in a vacuum the linkage through one circuit A due to unit steady current in a neighboring circuit B is equal to the linkage through B due to unit current in A; hence the mutual inductance of two circuits is the same, whichever is the primary circuit. If the circuits are closely coupled and have high self-inductance, and if an alternating electromotive force E be applied to A, the resulting ac induces an electromotive force in B approximately equal to $(N_B/N_A)\ E$; in which N_A and N_B represent the numbers of turns in the respective circuits. The principle is utilized in induction coils and in potential transformers.

Electromagnetic Field. Traditionally and in a simplistic way, it may be stated that a wire carrying an electric current is surrounded by a magnetic field whose lines of force are circles with the wire as their axis. This statement implies that the magnetic field is directly traceable to the moving electricity in the wire. This is an oversimplified explanation, however, because each electric "particle" projects into space a radiating field of electric force; and as the particles move along the wire, the lines of force move with them.

Theory of Maxwell. As will be developed later, the Maxwellian theory of the electromagnetic field has been supplanted and refined. Prior to the more detailed description of modern concepts, it may be in order to develop the earlier Maxwell concept. According to the theory of Maxwell, it is the motion of these lines of electric force that sets up the magnetic field transverse to them. More generally, a variable electric field is always accompanied by a magnetic field; and, conversely, a variable magnetic field is accompanied by an electric field. The joint interplay of electric and magnetic forces here described is what is called an electromagnetic field, and is considered as having its own objective existence in space apart from any electric charges or magnets with which it may be associated. An essential feature of the theory is that this process, whatever it is, represents a flow of energy at right angles to both electric and magnetic components. The flux density of this energy (corresponding to the intensity of radiation) is represented by what is known as the Poynting vector. Electromagnetic radiation is, on this theory, the propagation of these electric and magnetic stresses through space with the speed of light, somewhat as the much slower waves of elastic stress are propagated through steel. The conditions in an electromagnetic field are expressed mathematically by Maxwell's equations.

When an electric charge is set into motion, it builds about itself an electromagnetic field, and this implies a distribution of energy throughout space. The density of this energy at any point of the field is proportional to the product of the electric and magnetic vector components and the sine of the angle between them (vector product). The total field energy can be obtained by suitable integration, and is greater than that of the purely electric field of a stationary charge. Maxwell's theory treats this excess as kinetic energy, thus endowing the moving charge with an "electromagnetic mass" and an "electromagnetic momentum" inherent in its electrical character.

The Maxwell equations comprise a set of four classic formulae of the electromagnetic theory. They deal with certain vector quantities pertaining to any point of a region under varying electric and magnetic influence. If the point is in empty space, the equations are somewhat simplified; in general, provision must be made for the presence of dielectrics, conductors, or magnetizable bodies. In these equations, **H** is magnetizing force, **B** is magnetic induction, **E** is electric intensity, **D** is electric induction, ρ is electric charge density, **J** is conduction current density, t is time. The "curl" and the "divergence" of a function are well-known operators of vector analysis. The equations, in rationalized mks units are

$$\text{Curl } \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + \mathbf{J} \qquad \text{(I)}$$

$$\text{Curl } \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad \text{(II)}$$

The additional relations

$$\text{Div } \mathbf{B} = 0 \qquad \text{(III)}$$

$$\text{Div } \mathbf{D} = \rho \qquad \text{(IV)}$$

are frequently included as part of Maxwell's system, although they are not independent relations if one assumes the conservation of charge. The last two are also known as the Gauss law. For linear homogeneous isotropic media, $\mathbf{B} = \mu\mathbf{H}$; $\mathbf{D} = \epsilon\mathbf{E}$. The values of μ and ϵ for a vacuum satisfy

$$\mu_v \epsilon_v = 1/c^2$$

where c is the speed of light, μ is permeability and ϵ is permittivity.

Electromagnetic Radiation

The electromagnetic spectrum is the total range of wavelenghts or frequencies of electromagnetic radiation. As shown by the accompanying table, this range extends from the longest radio waves to the short cosmic rays.

ELECTROMAGNETIC SPECTRUM SHOWING PRINCIPAL RADIATION CATEGORIES

Frequency (Hz)	Type of Radiation	Wavelength (cm)	
10^{23}			
	Cosmic Rays	10^{-12}	
10^{22}			
		10^{-11}	
10^{21}			
	Gamma Rays	10^{-10}	
10^{20}			
		10^{-9}	
10^{19}			
	X-Rays	10^{-8}	
10^{18}			
		10^{-7}	
10^{17}			
		10^{-6}	
10^{16}	Ultraviolet Radiation		
		10^{-5}	
10^{15}	Visible Light		
		10^{-4}	
10^{14}			
	Infrared Radiation	10^{-3}	
10^{13}			
		10^{-2}	
10^{12}	Submillimeter Waves		
		10^{-1}	
10^{11}			
		1.0	
10^{10}	Microwaves (Radar)		
		10	
10^{9}			UHF
		10^{2}	
10^{8}	Television and FM Radio		VHF
		10^{3}	
10^{7}	Short Wave		HF
		10^{4}	
10^{6}	AM Radio		MF
		10^{5}	
10^{5}			LF
	Maritime Communications	10^{6}	
10^{4}			VLF

Prior to Maxwell's studies of the electromagnetic field in an effort to substitute electric and magnetic forces for elastic forces in the theory of light propagation, light was believed to be a transverse wave motion in an *ether* which behaved like an elastic solid. According to Maxwell's views, it was postulated that light is the result of vibrating electric charges. These set up alternating electric and magnetic fields at right angles to each other and to the direction of propagation, which pass on the energy from one portion of the ether to the next as an electromagnetic wave. (Poynting's theorem states that the rate of this energy transfer is proportional to the product of the electric and magnetic intensities.) The theory was successful in explaining many of the electrical and magnetic properties of light, such as the Faraday effect and the Kerr effect.

Maxwell suggested that it should be possible to produce waves of much longer wavelength than light by causing electricity to oscillate in a conductor. This was a forecast of the radio waves, upon which radio transmission depends, and which exhibit many of the characteristics of light, such as reflection, refraction, diffraction, interference, polarization, etc., but on a gross scale. Other researches showed that the infrared and ultraviolet radiations have these same properties. It was therefore natural to classify them together as the same phenomenon in different frequency ranges. X-rays for a time could not be identified with this group, but their diffraction by crystals finally demonstrated their wave character and they now take their place next to the ultraviolet. Meanwhile the quantum theory put a new aspect on the whole matter, and the gamma rays were added to the radiation family.

Present Concepts of Electromagnetic Theory. The task of electromagnetic theory is to account for the effects of electrical charges in various states of motion. Although historically electromagnetic theory was developed from Coulomb's celebrated law, it is at present more economic to develop it differently. The macroscopic effects are described with remarkable accuracy by the following set of equations (rationalized mks system of units):

$$\mathbf{F} = q\mathbf{E} + q\mathbf{v} \times \mathbf{B} \tag{1}$$

$$\nabla \cdot \mathbf{J} + \frac{\partial \rho}{\partial t} = 0 \tag{2}$$

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + \mathbf{J} \tag{3}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{4}$$

$$\mathbf{D} = f_1(\mathbf{E}) \tag{5}$$

$$\mathbf{B} = f_2(\mathbf{H}) \tag{6}$$

$$\mathbf{J} = f_3(\mathbf{E}, \mathbf{H}) \tag{7}$$

provided the functional relationships indicated in Equations (5), (6), and (7) are known explicitly. With these equations and the laws of mechanics, classical electromagnetic theory becomes essentially a branch of applied mathematics.

Equation (1), sometimes known as the Lorentz force equation, defines the field quantities, **E**, the electric field intensity, and **B**, the magnetic induction, in terms of an observable, the force **F** on a charge q. In Equation (1), **v** is the velocity of the charge relative to the observer. Equation (2) is a statement of the law of conservation of electric charge in terms of the charge density ρ and the total current density **J**. Equation (3) is the differential form of Ampère's law,

$$\oint_{\text{c of s}} \mathbf{H} \cdot d\mathbf{l} = \iint_S \mathbf{J} \cdot d\mathbf{S} = I$$

which relates the magnetic field intensity **H** to the current, including in addition the displacement current density term $\partial\mathbf{D}/\partial t$, which was added by Maxwell to make the law applicable to time-varying fields. The term **J** represents the total current density. Equation (4) is the differential form of Faraday's law of electromagnetic induction. Equations (5), (6), and (7) are functional relationships, for the most part determined experimentally, by means of which the effects of different materials are accounted for. Mathematically, these equations are employed to reduce Equations (3) and (4) to a pair of equations in only two unknowns. In free space, Equations (5), (6) and (7) take their simplest form, respectively, $\mathbf{D} = \epsilon_0 \mathbf{E}$, $\mathbf{B} = \mu_0\mathbf{H}$, $\mathbf{J} = 0$ (or $J = J_s$, a source current independent of **E** and **H**), where ϵ_0 and μ_0 are constants whose value depends on the system of units (in the mks system $\epsilon_0 = 8.854 \times 10^{-12}$, $\mu_0 = 4\pi \times 10^{-7}$). Since matter itself is a relatively dilute collection of charged particles, it is always theoretically possible to define terms so that the theory is a description of the effects and interactions of charges in free space, with consequently no essential distinction between **D** and **E** or between **B** and **H**, as indicated above. In practice, however, effects of materials are usually best handled in another way. Dielectric polarization effects are accounted for by making the **D** vector include the electric dipole moment density **P**, $\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}$, and then introducing a material constant, the permittivity ϵ, such that $\mathbf{D} = \epsilon\mathbf{E}$. The relative permittivity of a dielectric material is then equal to one plus the electric susceptibility. Magnetic polarization effects are handled similarly by defining the field vector **B** so that it includes the magnetic dipole moment density **M**, $\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M})$. The material permeability is then introduced so that it depends upon the magnetic susceptibility analogously, and $\mathbf{B} = \mu\mathbf{H}$. Effects of conductors are represented by a material conductivity σ, such that $\mathbf{J}_c = \sigma\mathbf{E}$. With these simple forms for Equations (5), (6) and (7), Equations (3) and (4) take on the useful form

$$\nabla \times \mathbf{H} = \epsilon\frac{\partial \mathbf{E}}{\partial t} + \sigma\mathbf{E} + \mathbf{J}_1 \tag{8}$$

$$\nabla + \mathbf{E} = -\mu\frac{\partial \mathbf{H}}{\partial t} \tag{9}$$

provided μ and ϵ are constant in time. The term $\mathbf{J}_1$ here includes currents arising from charges in free space plus any (source) currents which are independent of **E** and **H**. If there are no free charges in the region, $\mathbf{J}_1$ includes only the source currents; these latter are known, so Equations (8) and (9) may be solved for **E** and **H**. Since the equations are partial differential equations, boundary conditions over closed surfaces are required for unique solutions. Boundary conditions on the field quantities, which must hold at any boundary between two regions, may be derived from these equations. The conditions are: across a boundary (a) tangential **E** must be continuous, (b) tangential **H** must be continuous, (c) normal **D** and normal **B** must be continuous. Idealizations of material properties are sometimes helpful. For example, a perfect conductor has no nonstatic fields inside it, and, at its surface, tangential **E** and normal **B** are zero, tangential **H** is equal and perpendicular to any surface current density, and normal **D** is equal to any surface charge density.

Two additional equations, especially useful in static problems, may be deduced from Equations (2), (3) and (4):

$$\nabla \cdot \mathbf{D} = \rho \tag{10}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{11}$$

Solutions to the field equations are most readily obtained by imposing a restriction on the time dependence. If the fields are assumed to be independent of time (static), then Equations (3) and (4) or (8) and (9) decouple. One of the equations becomes $\nabla \times \mathbf{E} = 0$. This means that **E** is irrotational and may be represented by a scalar potential function, ϕ, $\mathbf{E} = -\nabla\phi$. Combining this with Equation (10) gives the fundamental equation of electrostatics,

$$\nabla^2\phi = -\rho/\epsilon \tag{12}$$

Poisson's equation. This equation for the electrostatic potential is solved by the standard methods of partial differential equations. The boundary conditions on the potential may be found from the boundary condition on the fields. In practice, it is frequently necessary to solve for the potential and electric field in a restricted region in which the charge density is zero, but the potential at the boundary is held at some particular value(s). The problem then is to solve Laplace's equation, $\nabla^2\phi = 0$, subject to the stated boundary conditions. The standard techniques for solving boundary value problems are employed. However, if the region of interest is partially open, known analytical techniques are sometimes inadequate to solve the problem. In two-dimensional problems of such a difficult type, the method of conformal transformations (conjugate functions) is often helpful.

The main applications of electrostatic theory are in (a) the theory of material properties, (b) the calculation of charged particle trajectories in electron guns, deflection systems, and accelerators (here in conjunction with magnetostatic theory), (c) the calculation of circuit component values, such as capacitance, and (d) the determination of voltage gradients in connection with voltage breakdown problems.

Magnetostatic theory is developed from Equations (11) and (8). Since **B** is divergenceless, it can be represented by the curl of a vector **A**, which is known as the magnetic vector potential. Equation (8) can usually be written in terms of this potential as follows:

$$\nabla^2\mathbf{A} = -\mu\mathbf{J} \tag{13}$$

Taken one rectangular component at a time, this equation is of the same form as Poisson's equation [Equation (12)] and may be solved in the same way. The boundary conditions on **A** may be found from those on **B** and **H**. In regions with no current, Equation (8) becomes $\nabla \times \mathbf{H} = 0$ so that **H** may be represented by a scalar potential function $\mathbf{H} = -\nabla\phi_m$. In such regions then, in view of Equation (11), the magnetic scalar potential, ϕ_m, must satisfy Laplace's equation

$$\nabla^2\phi_m = 0 \tag{14}$$

provided $\nabla\mu = 0$ in the region. The techniques and solutions of electrostatics are applicable to many magnetostatic problems. Unfortunately, however, in practice many of the systems designed to establish a given magnetic field incorporate ferromagnetic materials. For such materials, the magnetic susceptibility (and hence the permeability) is not independent of the field intensity and the field equations become nonlinear. Present mathematical techniques for handling nonlinear problems are severely limited. Practical magnetostatic problems are, therefore, frequently solved by some approximation. One of the simplest and most useful approximations is a representation by a magnetic circuit. Series and parallel branches of the magnetic circuit may be recognized, and the techniques of linear and nonlinear circuit analysis can be applied to obtain a solution.

Magnetostatic theory is applicable to a myriad of magnetic devices including deflection systems, motors, generators, relays, magnetic pickup devices, permanent magnets, memories, transducers and coils. To date, the need for particular solutions has frequently arisen before sound analytical methods have been available, so many devices are developed empirically.

Energy is required to establish electric and magnetic fields, and such energy is associated with the fields. The field energy in a given volume may be computed in most cases from a volume integral of one or both of the following energy density expressions: $W_e = \frac{1}{2}\epsilon E^2$, $W_m = \frac{1}{2}\mu H^2$, respectively the electrostatic and magnetostatic values.

When the fields are time varying, Equations (8) and (9) are coupled and must be solved simultaneously. Almost invariably, a potential function such as a vector potential or a Hertz potential is introduced. For example, Equation (11) implies that **B** may be replaced by a vector potential such that $\mathbf{B} = \nabla \times \mathbf{A}$. Equation (9) implies the following equation for **E**,

$$\mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t} \tag{15}$$

so that **H** and **E** may be replaced in Equation (8), and with the condition on **A**, $\nabla \cdot \mathbf{A} = \mu\epsilon\, \partial\phi/\partial t$, the following equations may be obtained for **A** and ϕ (σ assumed zero here):

$$\nabla^2\mathbf{A} - \mu\epsilon\frac{\partial^2\mathbf{A}}{\partial t^2} = -\mu\mathbf{J} \tag{16}$$

$$\nabla^2\phi - \mu\epsilon\frac{\partial^2\phi}{\partial t^2} = -\rho/\epsilon \tag{17}$$

That is, both the vector potential **A** and the scalar potential ϕ satisfy a differential equation known as the inhomogeneous wave equation.

Because of their simplicity and practical importance, solutions for those sources and fields which simply oscillate at a single frequency have been studied extensively. In this case, the time is eliminated as an independent variable, as if by a transform operation. (In fact, transform methods are often the best means of obtaining transient field solutions.) In the equations, the time derivatives are replaced by frequency multipliers so that the resulting equations are functions of the space variables only. The vector potential may then be found by standard techniques of partial differential equations and boundary value problems. Having **A,** the field quantity **B** is found from $\mathbf{B} = \nabla \times \mathbf{A}$ and **E** is found from Equation (8). In practice, a theorem which can be derived from the field equations, called the reciprocity theorem, is often helpful. The theorem relates the fields $\mathbf{E}_a$ and $\mathbf{E}_b$ produced respectively by a pair of current distributions $\mathbf{J}_a$ and $\mathbf{J}_b$. The theorem is

$$\iiint \mathbf{E}_a \cdot \mathbf{J}_b dv = \iiint \mathbf{E}_b \cdot \mathbf{J}_a dv \tag{18}$$

For example, if $\mathbf{J}_b$ is selected to be a point current at point P, directed along x (represented mathematically by a Dirac delta function), then Equation (18), $E_{ax}(P) = \iiint \mathbf{E}_b \cdot \mathbf{J}_a\, dv$, gives a formula for the computation of the field due to $\mathbf{J}_a$ which is equivalent to a superposition integral involving a Greens function.

Perhaps the most fundamental problem of electromagnetic theory is the determination of the fields of a point charge, at rest, in oscillation, or in some general state of motion. For a point charge q, at rest in free space, the solution may be obtained by solving Equation (12) in spherical coordinates. With the point charge at the origin, symmetry conditions may be employed to eliminate the angular variation, and the remaining differential equation in r can be solved subject to Equation (10) to give $\phi_G = (q/4\pi\epsilon_0 r)$ for the potential associated with the point charge. A superposition integral

$$\phi = \iiint \frac{\rho\, dv}{4\pi\epsilon_0 r} \tag{19}$$

may then be employed to find the potentials associated with more complicated distributions. The field of an oscillating dipole, which is equivalent to a point alternating current, is also of great interest. This solution may be obtained from Equation (16) (single frequency version). If the point current is directed along z, the z-component of the vector potential may be found by a procedure similar to that employed for a point charge. The final result is

$$A_{ZG} = \frac{I\, \Delta Z}{4\pi\mu_0 r} \cos\omega(t - \sqrt{\mu_0\epsilon_0}\, r) \tag{20}$$

where $\mathbf{I}\, \Delta z$, the current moment, is equal to $q\, \Delta z$, the maximum dipole moment of the oscillating dipole. The factor $(t - \sqrt{\mu_0\epsilon_0}\, r)$ exhibits the time delay required for the effects of the oscillating charges to propagate to distant points. The electric and magnetic fields may be computed from Equation (20) as indicated above. The magnetic field strength produced by an oscillating dipole (point current) is, for example, in the spherical coordinate system (r, θ, φ)

$$H_\varphi = \frac{I\, \Delta Z}{4\pi} \sin\theta \left[\frac{\cos\omega(t - \sqrt{\mu_0\epsilon_0}\, r)}{r_2} - \frac{\omega\sqrt{\mu_0\epsilon_0}}{r} \sin\omega(t - \sqrt{\mu_0\epsilon_0}\, r) \right]$$

This form, like Equation (20), shows that the crests and valleys of the field oscillations are propagated in spherical waves at the speed of light $v = (\mu_0\epsilon_0)^{-1/2}$. The solution for a point current may be employed in an integral similar to Equation (19) to find the vector potential of a more complicated distribution of current. Such solutions may also be employed to find the radiation patterns and input impedances of antennas.

The potentials and fields produced by a charge moving in an arbitrary way may also be obtained.

In regions free of source currents and charges, the fields and potential satisfy the homogeneous wave equation [for example Equation (16) with $\mathbf{J} = 0$]. Then one of the simpler solutions which can be obtained is that of the plane electromagnetic wave. With appropriate orientation of the rectangular coordinate system, the solutions show that plane waves may progress along z, with components as follows:

$$E_x = E_0 \cos\omega(t - \sqrt{\mu_0\epsilon_0}\, z)$$

$$H_y = E_0 \sqrt{\frac{\epsilon_0}{\mu_0}} \cos\omega(t - \sqrt{\mu_0\epsilon_0}\, z)$$

where E_0 is an arbitrary constant amplitude. Note that **E, H** and the direction of propagation are all perpendicular to one another. The Poynting vector, $\mathbf{S} = \mathbf{E} \times \mathbf{H}$, points in the direction of the propagation. Moreover, the power carried through a closed surface by an electromagnetic field may be computed from a surface integral of the Poynting vector.

With single frequency fields in source free regions, both **H** and **E** can be represented by vector potentials, $\mathbf{H}_1 = \nabla \times \mathbf{A}_1$, $\mathbf{E}_2 = \nabla \times \mathbf{A}_2$, and moreover the coordinate systems may be oriented so that $\mathbf{A}_1$ and $\mathbf{A}_2$ each have a single component. In cylindrical systems, this single component is commonly along z. $\mathbf{H}_1$ is then transverse to z (TM) the set of fields, $\mathbf{E}_1$, $\mathbf{H}_1$, derivable from $\mathbf{A}_1'$, are called TM fields. $\mathbf{E}_2$is likewise transverse to z and the set of fields, $\mathbf{E}_2$, $\mathbf{H}_2$, derivable from $\mathbf{A}_2$, are called TE fields. This procedure is particularly helpful in problems involving transmission lines and waveguides.

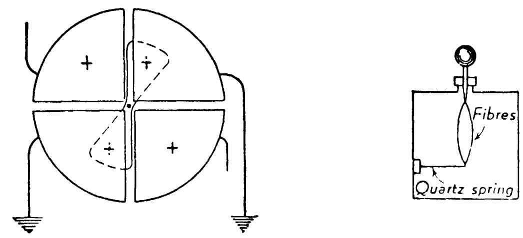
Some of the most interesting and fundamental problems of electromagnetic theory are concerned with the scattering and diffraction of electromagnetic waves. For example, exact solutions are available for the scattering by cylinders and spheres, as well as an infinitely long slit. Approximate solutions are available for many other shapes. The methods are those outlined above, supplemented by generalizations of the principles of Huygens and Babinet.

Another topic of wide interest is the nature of fields in ionized gases or plasmas. The applications range from ionospheric propagation to microwave devices to nuclear apparatus to magnetohydrodynamics to satellite reentry problems. The simplest theory for these effects is developed from Equations (3) and (4) (single frequency version) by separating the ion current term $\mathbf{J}_e = \rho\mathbf{v}$ from **J,** and employing Newton's law to eliminate **v** in favor of **E, H** and whatever mechanical constraints are applicable.

Effects peculiar to charges moving with very high velocities have not been included in this discussion. Numerous entries throughout this encyclopedia are concerned with electromagnetic phenomena. See also **Magnetism.**

ELECTROMETER. An instrument for measuring electric charge, usually by mechanical forces exerted on a charged electrode in an electric field. It consists, therefore, of a sensitive voltmeter operating on the principles of electrostatic attraction and repulsion. Thus, if the movement of the gold leaf in an electroscope is observed through a microscope whose ocular is provided with a calibrated scale, the instrument becomes an electrometer, capable of measuring potential differences in millivolts. (Some forms of electrostatic voltmeters operate on the same principle.) If the capacitance of the charged system is known, the rate of movement of the electrometer index may be used to measure the current from the discharging body; ionization currents are often thus measured.

The quadrant electrometer has a thin, oblong, metal plate suspended horizontally in the interior of a flat, circular metal box cut into four quadrants. One pair of opposite quadrants and the suspended strip are connected to the source of potential, the other pair of quadrants is grounded. This causes the strip to turn toward the grounded pair against the torsion of the suspending wire (see figure, *left*). Several electrometers have been designed, depending upon the lateral deflection of a lightly stretched, silvered or platinized quartz fiber in an electric field; they are called string electrometers. The Wulf electrometer employs two such fibers side by side; on being charged, they bulge apart (see figure, *right*). The displacement of the fibers is observed in a micrometer mi-



Electrometers: (*Left*) Quadrant electrometer showing quadrants. (*Right*) Quartz fiber electrometer.

croscope. Some of the special electrometers used for work with cosmic rays are of this type.

Vacuum tube electrometers, which are specially designed amplifiers with high input impedance, have been replaced by quadrant and string electrometers to a large extent.

See also **Electrical Instruments;** and **Electroscope.**

ELECTROMOTIVE FORCE. The electric potential difference (emf) between the terminals of any device which is used or may be used as a source of electrical energy, i.e., to supply an electric current. More strictly, the limiting value of that potential difference which is found as the current flowing from the source approaches zero. To avoid ambiguity, the strict sense of the term is often indicated by the use of the qualifying term "open circuit" or "no load."

The open circuit electromotive force of a cell is identical with its reversible potential difference; that of a rotating electrical machine is the potential difference existing across its terminals when the machine is neither receiving nor delivering electric power, i.e., is at the transition point between being a generator and being a motor. When a potential source of electrical energy, such as a capacitor, an inductor, or a rotating machine, is receiving energy from the external circuit, it is said to develop a counter-electromotive force. See also **Kirchoff Laws of Networks.**

Counter electromotive force is the emf generated by a running motor, by virtue of its generator behavior, or by an inductive circuit element through which the current is increasing with time. The total emf in the circuit is the impressed emf minus the counter emf; the current is given by the ratio of this total emf to the resistance in the circuit. The counter-electromotive force is sometimes called the back electromotive force. The *impressed emf* is the open-circuit (no load) emf of a source connected into a network.

Effective or root-mean-square electromotive force is the effective value of an alternating electromotive force and is the square root of the mean value of the squares of the instantaneous values. Where the variation of voltage with time can be expressed by a mathematical function, the value of the effective current can commonly be expressed in terms of the maximum value of the emf. Thus, for a sine wave relationship,

$$E_{eff} = \frac{E_{max}}{\sqrt{2}}$$

ELECTROMOTIVE SERIES. See **Activity Series.**

ELECTROMYOGRAPHY. The recording of action potentials from muscles in the living subject by means of a needle electrode inserted through the skin into the muscle and connected to a suitable amplifier and cathode-ray oscillograph. Normal resting muscle shows no changes in potential; contraction gives rise to large numbers of monophasic or diphasic spikes, indicating changes in potential of up to 1 millivolt, each lasting from 5–10 milliseconds. Various diseases and injuries of muscles and their nerves of supply give rise to characteristic and distinct departures from these normal patterns, and hence electromyography is of value in assessing the extent of damage in diseases of the nervous system in which muscle is involved and in estimating the probability, progress and degree of recovery.

ELECTRON. An elementary particle of rest mass

$$m = 9.107 \times 10^{-31}$$

kilogram, a charge of 1.602×10^{-19} coulomb, and a spin quantum number $\frac{1}{2}$. Its charge may be positive or negative, although the term electron is commonly used for the negative particle, which is also called the negatron. The positive electron is called the positron. The electron is a constituent of all matter, thus the normal atom consists of a positively charged nucleus surrounded by a sufficient number of electrons so that their total charge is equal to the positive charge on the nucleus. The electron also has wave characteristics, with a frequency and a wavelength. According to wave mechanics, an electron traveling with speed v is associated with a "de Broglie wave" train of wavelength $\lambda = h/mv$ (in which m is the electronic mass and h is Planck's constant).

Bonding Electron. An electron in a molecule which serves to hold two adjacent nuclei together.

Conduction Electron. An electron which plays an important part in electrical or thermal conduction by solids, i.e., by metals or semiconductors, e.g., the electrons in the conduction band, which are free to move under the influence of an electric field.

Electron Donor. 1. When a valence bond between two atoms is that type of covalent linkage in which both the electrons of the duplet are supplied by one atom, then that atom, or portion of the molecule of which it forms a part, is called the electron donor, and the other atom in the linkage is called the electron acceptor. 2. A donor is also an impurity added to a pure semiconductor to increase the number of free electrons.

Electron Duplet. A pair of electrons which is shared by two atoms, and is equivalent to a single, nonpolar chemical bond.

Electron Octet. A group of eight valence electrons which constitutes the most stable configuration of the outermost, or valency, electron-shell of the atom, and hence the form which frequently results from electron transfer or sharing between two atoms in the course of a chemical reaction.

Electron Pair. The negatron and positron that result from the pair-production process or interact to initiate an annihilation process.

Electron Shell. The structure of a neutral atom consists of a positively charged nucleus with a number of electrons moving about it—the number being such that their total negative charge is equal to the positive charge on the nucleus. These electrons may be assigned to various shells, characterized by different principal quantum numbers.

Electron Spin. The intrinsic angular momentum of an electron, independent of any orbital motion. Spin ($= h/2$) contributes to the total angular momentum of the electron and is quantized. It gives rise to multiplicity in line spectra, which may be characterized by introduction of the spin quantum number.

Electron Transfer. The process of the shifting of an electron from one electrical field to another, as in the formation of an electrovalent bond, in which an electron moving in an orbit about one atom shifts to move in an orbit around the two bonded atoms.

Free Electron. An electron which is not restrained to remain in the immediate neighborhood of an atom or molecule.

Orbital Electron. An electron remaining with a high degree of probability in the immediate neighborhood of a nucleus, where it occupies a quantized orbital.

Photo Electron. An electron ejected from a substance by the action of a single photon of light or other electromagnetic radiation.

Secondary Electron. An electron deriving its motion from a transfer of momentum from primary radiation, which may be either particulate or electromagnetic.

Valence Electron. The electrons in the outermost shell of the structure of an atom. Since these electrons are commonly the means by which the atom enters into chemical combinations—either by giving them up, or by adding others to their shell, or by sharing electrons in this shell—these outermost electrons are called valence electrons.

ELECTRON AFFINITY. 1. Degree of electronegativity, or the extent to which an atom holds valence electrons in its immediate neighborhood, compared to other atoms of the molecule. 2. The work required to remove an electron from a negative ion, and hence to restore the neutrality of an atom or molecule, is called the electron affinity of the atom or molecule.

ELECTRON BEAM. A stream of electrons moving with about the same velocity and in the same direction, so as to form a beam.

ELECTRON BEAM LITHOGRAPHY. In chip making, the processing cost is only indirectly related to the amount of circuitry on a chip and thus it is an economic advantage for the semiconductor industry to increase the amount of circuitry per chip. This can be accomplished in either of two ways: (1) increasing the size of the chip, or (2) reducing the size of the circuits themselves. The latter approach which has other advantages in terms of circuit performance and the application of chips by electronic equipment manufacturers, commercial and military, has predominated integrated circuit technology, with a target for the year 2000 or earlier of one billion transistors on a single chip. In passing, it should also be pointed out that larger chips exhibit poorer manufacturing yields and less mechanical stability. Hence, increasing circuit density can best be accomplished by reducing the circuit dimensions. There are physical limitations, of course, to the width of lines and separations in an intensely crowded chip. Over the last couple of decades, these dimensions rapidly dropped from the millimeter dimensions of the very early days of solid state electronics to micrometers (microns) as circuits proceeded from medium- to large- to very-large-scale integration—to the point as of the late 1980s where submicron dimensions are impacting the industry.

As pointed out in the article on **Microstructure Fabrication (Electronics)**, earlier optical lithography methods became limited by diffraction and optical aberrations, a situation which occurs at feature sizes below about 2 micrometers.

Electron beam lithography entered the scene in the late 1970s as a means to fill the gap.

Principles of Electron Beam Lithography. Typically, an electron beam lithography system accelerates and focuses an intense beam of electrons to draw precise circuit patterns on suitable substances. Electron beam instruments, unlike optical instruments, require a high-vacuum environment for both the electron beam path and the substrate.

An intense beam of electrons from a high brightness source, such as lanthanum hexaboride, is focused to a size of about 0.2 micrometer or less by electromagnetic lenses. Patterns are generated by moving the beam across the substrate and turning the beam on and off with a blanking system. Two methods are employed to move the beam: (1) small movements are accomplished by deflecting the beam with electromagnetic coils; (2) large movements are made by moving the wafer by a mechanical stage. Positioning precision is achieved in the first case by using highstability electronics of high numerical precision, while in the second case it is achieved by employing a laser interferometer to monitor the stage motion. Because the beam must move at rates of two to ten million steps per second, a computer is required to control the position. Few computers are capable of the necessary data rates and, therefore, specific custom hardware is usually included in the interface to reduce the required amount of data.

The technique used in writing with an electron beam is similar to that used for tracing figures on cathode ray tubes in computer graphics displays. Two distinct methods have been developed for this purpose: (1) raster scan, and (2) vector scan (sometimes called *random draw*). Because specific and distinct hardware is required for each, electron beam instruments are designed to use either one method or the other, but not both.

Raster Scan Method. In the raster scan method, the beam is moved across the entire chip using a television-type raster. It is unblanked only where exposure is desired. One system moves the stage continuously in one direction, *X*, while scanning the beam in an orthogonal direction, *Y*. In this way, a narrow strip is exposed, and the entire wafer is covered by exposing a large number of these strips. As the beam is scanned in the *Y* direction, the hardware must turn it on or off at each point. Currently available computers are not capable of supplying data at the required high rates and, therefore, a large block of information is stored in the hardware and used repeatedly to expose each of many identical areas on the wafer. This method results in fast exposure, but only if many identical chips are to be exposed. The raster scan method is shown in Fig. 1.

Advantages of the raster scan method include less stringent requirements on the electronics and on the electron optics. In some cases, it is possible to achieve faster exposures. Deflection of the beam in the *X* direction is limited to that necessary to correct for minor errors in the position of the stage. In the *Y* direction, the beam deflection required is significantly less than that for the vector scan method. At these small

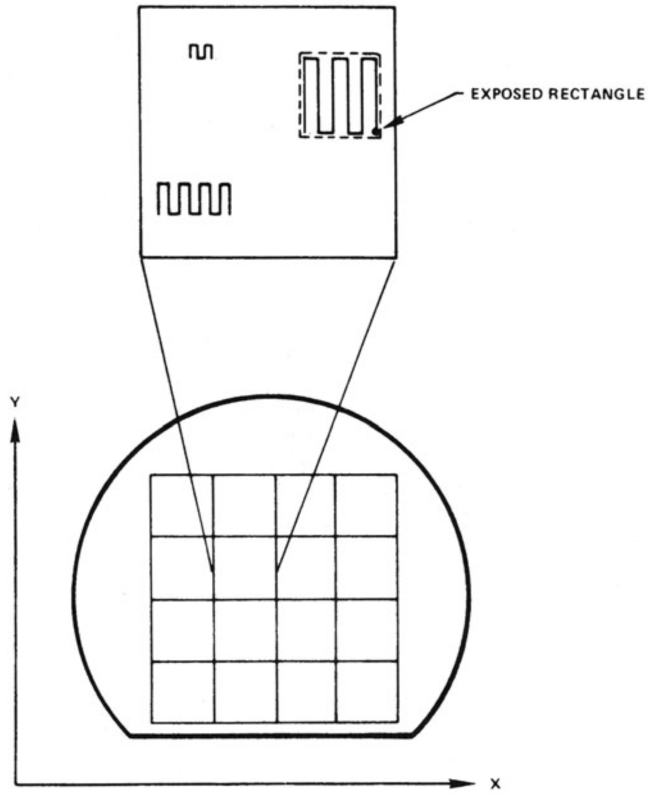


Fig. 1. Exposure of wafer by raster scan method.

beam deflections, field distortion is normally not a problem. Processing times are short if there are many identical chips to be made from a single wafer, but increase as the size of the chip increases.

These advantages are offset by a number of disadvantages. The necessity of "stitching" together many small fields places stringent requirements on the precision of the mechanical stage. The entire wafer must be scanned, and this is time consuming if it is necessary to expose only a small fraction of the area. Furthermore, corrections for the exposure effects of adjacent fields and for wafer distortion are difficult, if indeed possible.

Vector Scan Method. In the vector scan method, the beam is moved only to locations where exposure is desired. The beam is moved over a serpentine path to cover an entire basic element, the portion of the chip which can be exposed by a single computer command. The element is constructed of a series of parallel lines, each line consisting of a row of exposed points. Because the spacing of the points is of the same order as the width of the beam, the entire area is exposed. The maximum point rate is 5 or 10 megahertz, and, because most computers are unable to give commands at this rate, hardware is normally included to move the beam over an entire clement. The beam is blanked only for a short time between the tracing of adjacent lines and while moving from one element to another. The vector scan method is shown in Fig. 2.

The vector scan method has the advantage of not using up time moving the beam over areas which are not to be exposed. In addition, it is relatively easy to make corrections for scan field distortions, for the effect of adjacent exposed areas, and for wafer distortion. The necessary size of the scan field, however, is significantly larger than that used in the raster scan method. Consequently, more stringent requirements are placed on the quality of the electron optics, the scan coils, and the amplifier which drives them.

A computer is used to drive an interface which implements the vector scan method. The interface contains a series of electronic counters and digital-to-analog converters which determine the position of the beam. The output from this circuitry is fed to an amplifier which drives the deflection coils. See Fig. 3.

Each pattern consists of a group of geometrical elements. The hardware stores the information which defines the coordinates of the lower

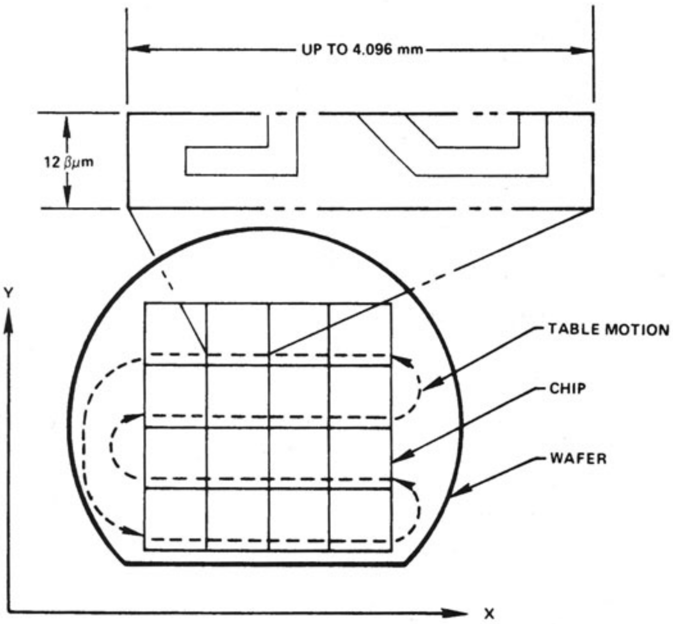


Fig. 2. Exposure of wafer by vector scan method.

left corner of each element also its height and its length. The hardware counts the height vector either up or down to expose a single line and decrements the length vector between lines to expose the entire element. The most commonly used element is a rectangle, but the hardware has the additional capability of drawing parallelograms and triangles.

The circuit designer defines the patterns, using another computer, usually one with graphics capabilities. The pattern definition is then placed on a magnetic tape which can be read by the machine's computer. The software in the latter machine then interprets the tape and produces the commands to drive the instrument. This software is capable of accepting control information related to scaling, exposure, and position, either from the magnetic tape or directly from the operator.

The customizing of the system maximizes flexibility in order to facilitate the engineering prototyping of circuits. A single wafer may contain either many identical or many different chips, and exposure and scale may be varied from chip to chip. The reduced number of steps in the process, as compared with optical methods, reduces the time necessary to make a modification and to obtain a finished circuit ready for testing. The result is a shorter turnaround time in the design of new circuits.

Fabrication Techniques. Electron beam techniques of fabricating devices are in many ways similar to conventional optical techniques. Processes common to both include growing and removing oxides, depositing metal conductors, and doping the semiconductor. Each process

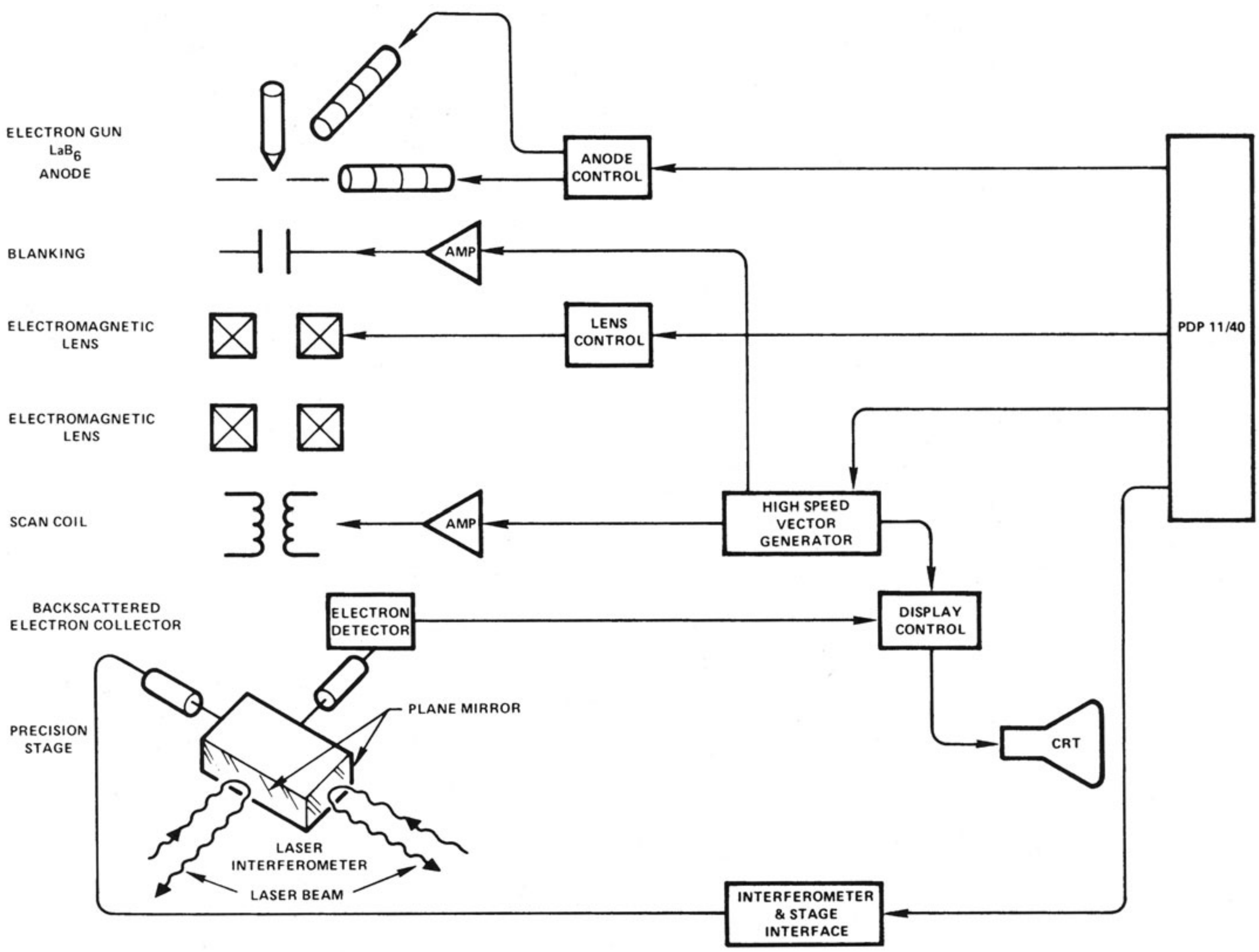


Fig. 3. Early 1980s electron beam instrument. (*GTE Labs.*)

requires that a different pattern be exposed and positioned accurately (registered) relative to previous ones.

Traditional methods of device fabrication define these pattern by a set of masks fabricated by optical techniques, usually using more than one processing step to manufacture each mask. The masks are then used to expose a photosensitive emulsion on the wafer, again by optical techniques.

Electron beam techniques make possible two improved methods of fabrication. (1) In the first, the electron beam instrument is used to make a mask. The masks are then used to expose the resist on the wafer with radiation. (2) The second electron beam technique is direct fabrication. Here the wafer itself is coated with an electron-sensitive polymer (resist), which is exposed directly by the electron beam. The wafer is then removed from the instrument, the resist developed, and the wafer processed. Each processing step requires that this sequence be repeated. In some cases, direct fabrication increases the processing time, but the technique has significant advantages, which include flexibility, greater ease of registration, and higher resolution. The devices commonly fabricated by direct electron beam methods have smaller dimensions and, therefore, require greater positioning accuracy. The accuracy is achieved by the use of registration marks consisting of heavy metal deposits or of holes in an oxide layer. The instrument uses electrons scattered from these marks to locate accurately the position of previous patterns. This technique achieves a positioning precision of 0.1 micrometer.

The process used to fabricate registration marks is similar to that used to fabricate the pattern. The device is covered first with a polymeric material and then with a thin aluminum coating which is used to drain charge from the device. Irradiation increases the degree of polymerization in some polymers and it reduces it in others. Chemicals are then used to develop the pattern by removing the unpolymerized material. The remaining polymer is used as a mask to protect areas not to be covered with metal, or not to be doped. Finally, the polymer is dissolved, and any excess metal or dopant is removed. This method is known as the "lift off" technique and is shown in Fig. 4.

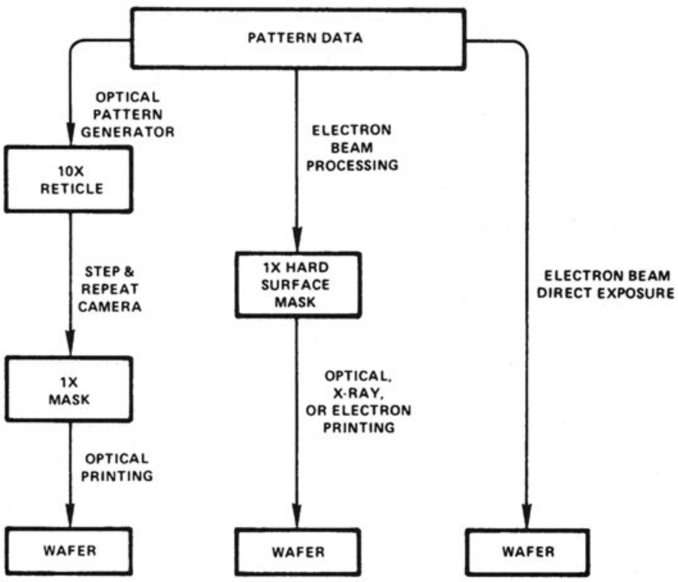


Fig. 4. Comparison of optical and electron beam fabrication.

The polymeric material used can be either of two general types: (1) In the first, positive resist, the exposed portion is depolymerized and therefore removed by the developer; while (2) in the second type, negative resist, the unexposed area is removed. The electrical charge per unit area (coulombs per square centimeter) that is required to depolymerize or to polymerize the resist is a definition of its sensitivity. The resolution of the resist describes the minimum feature size that can be usefully exposed. Poly(methylmethacrylate), PMMA, is an example of a well-known positive resist with high resolution, but poor sensitivity. Poly(butene-l-sulfone), PBS, and Poly(glycidylmethacrylate), PGMA, are well-known negative resists with higher sensitivity, but less resolution than PMMA. Resolution is degraded by electron scattering from the resist and from the substrate. Because the area exposed by the scattered electrons is much larger than the minimum area of the electron beam, the scattering is the limiting factor in determining the minimum possible feature size.

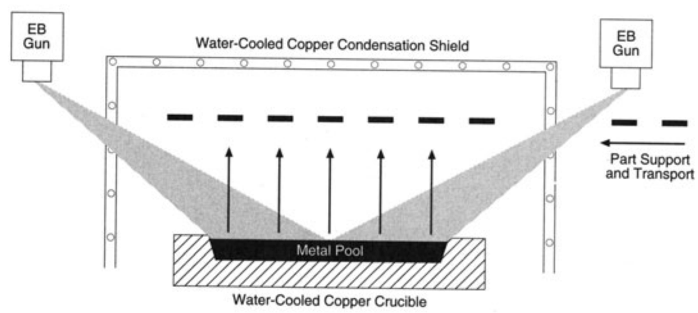
Because of the complexity of electron beam instruments, a substantial investment in equipment is required. These instruments, however, provide increased flexibility and automated control which are extremely useful in the design and prototyping of circuits.

Adapted from an article by W. D. Jensen, F. B. Gerhard, Jr., and D. M. Koffman, GTE Laboratories, Waltham, Massachusetts for *GTE Profile*

ELECTRON BEAM WELDING. See **Welding.**

ELECTRON BEAM VACUUM-EVAPORATION PROCESS. For decades, various coating processes have been developed to provide metal parts, fundamentally selected for their strength, with protection required to resist corrosion and to provide additional performance characteristics, such as tolerance of high surface temperatures. Frequently, gas turbines are cited as an example of such needs. High-temperature processes, such as dipping (zinc galvanizing, cadmium coating, etc.) have been used for steel pipes and fittings, and numerous articles have been electroplated with nickel, chromium, gold, silver, etc., to deter corrosion and tarnishing. Because of extremely demanding operating conditions, however, traditional coating processes no longer suffice. The gas turbine blade is an example, but this problem also arises in other industries, notably in the manufacture of electronic components, where thin films must be created.

The need for coating purity and in film formation has become extremely demanding. Where alloy components of coatings, for example, must be melted and vaporized, conventional heating processes introduce undesired impurities. Electron beam heating under vacuum creates the needed melting and vaporizing temperatures without introducing any impurities. See accompanying figure.



Sectional view of electron-beam vacuum-evaporation process. Desired alloy ingots are melted in a copper crucible by energy from two focused electron-beam guns, causing metal vapors to rise and coat parts in the upper part of a water-cooled chamber. Nearly all types of materials can be evaporated. For parts which require corrosion resistance at high-temperature operating conditions (such as gas-turbine blades), numerous alloy components (such as nickel, cobalt, chromium, aluminum, and yttrium among others) can be deposited on parts. The chamber is evacuated, parts are preheated, and the time of part exposure is in the five-minute time range. Films of 0.1 to 0.2-millimeter (0.004 to 0.008 inch) are deposited on the parts.

A typical batch-coating cycle for coating aircraft gas-turbine blades consists of:

- Loading blades into the processing chamber. During the coating process—wherein such alloy metals as chromium, aluminum, and yttrium form a coating of a basic material such as iron, cobalt, and nickel—the parts are manipulated to orient all surfaces so that a uniform coating can be maintained. This operation, of course requires automation under vacuum conditions.
- The chamber is evacuated to a pressure of about 2 Pa (0.02 mbar). Approximately five minutes are required to reach this condition.

- Parts must be preheated to 960–980°C (1760–1795°F), depending on the part alloy. Preheating is carefully controlled so that the condensing vapor particles are sufficiently mobile to promote the formation of compact grain structures and, consequently, dense films.
- In an average situation, a 0.1 to 0.2 mm (0.004 to 0.008 inch) thick film is deposited on the parts. This requires about 10 minutes. Depending on how the evaporation sources are arranged, the coating chamber will contain two or three electron beam guns, each with a beam-power rating of 200 KW max.
- Coated parts are returned to a load-lock chamber and allowed to cool below 300°C (570±F). The locked chamber is vented to atmospheric pressure and parts are unloaded.

As pointed out be Lämmermann and Kienel, "The composition and microstructure of the deposited film are the two major factors that determine its corrosion resistance. The typical composition of a MCrAly film is 20 Cr, 10 Al, 0.3 Y, ball M (Fe, Co, or Co-Ni). In a typical situation, the chromium, aluminum and yttrium contents of a Ni-CoCrAly deposit are 80%, 60%, and 60%, respectively, of acceptable tolerance ranges. Similarly, the cobalt, chromium, aluminum, and yttrium contents of a CoCrAly deposit are 60%, 45%, 60%, and 60%, respectively.

The hot-corrosion (and oxidation) resistance of coated parts can be improved further by applying a layer of thermal insulation. This coating must be sufficiently thick, have a low thermal conductivity, and hight thermal-shock resistance. Internal voids in the thermal coating also are helpful and can reduce the thermal conductivity of the material to a value well below that of the bulk material. The temperature difference between the outer surface of the coating and the outer surface of the underlying corrosion-resistant material can be as high as 150°C (260°F). In addition to reducing the temperature at the surface of the superalloy, the coatings also reduce thermal-shock loads on the parts, and rapid changes in ambient temperature are moderated and attenuated before they reach the substrate.

Zirconium oxide (ZrO_2) stablized by an addition of about 7% yttrium oxide (Y_2O_3) can be quite effective. Any oxygen deficit due to partial dissociation is compensated for by adding metered amounts of oxygen during the coating process. The crucibles used are considerably smaller than those used for evaporating metals. The evaporation temperatures of ZrO_2 and Y_2O_3 are so high that the EB guns are not sufficiently powerful to maintain a uniform temperature at the surface of a large-area melt pool.

Thermal spraying in the form of low-pressure plasma-spraying (LPPS) is another method for providing the high coating purity and strong coating-to-substrate bond needed by some products. See also *Thermal Spraying*.

Additional Reading

Lämmerman, H., and G. Kienel: "PVD Coatings for Aircraft Turbine Blades," *Advanced Materials & Processes*, 18 (December 1991).

Schilke, P. W., et al.: "Advanced Materials Propel Progress in Land-Based Turbines," *Advanced Materials & Processes*, 22 (April 1992).

ELECTRON DIFFRACTION. Beams of high-speed electrons exhibit diffraction phenomena analogous to those obtained with light, thus showing the wave-like character of electron beams. Such patterns are useful in the interpretation of the structure of matter.

See also **Electron Microscope.**

ELECTRONEGATIVITY. This term refers to the relative tendency of an atom to acquire negative charge. This tendency is not precisely defined because no exact theoretical or experimental method of evaluation has been devised. Electronegativity exists because the electronic cloud which surrounds each atomic nucleus is inadequate to block off the nuclear charge completely at the periphery. In other words, although every complete atom is electrically neutral as observable from a distance, a small fraction of the total nuclear positive charge can be detected at any point near the surface of the atom. This "effective nuclear charge" at the surface is relatively insignificant in the absence of low energy vacancies capable of accommodating a foreign electron, as in the atoms of M 8 elements (commonly called the *inert* or *noble gases*). However, wherever a low energy electron vacancy occurs in the outer shell of an atom's electronic cloud, the effective nuclear charge as manifest within that vacancy is of major importance. Indeed, it constitutes the cause and means of chemical bond formation, and largely determines both polarity and strength of the bond. Electronegativity is a measure of the force of this effective nuclear charge within an orbital vacancy at the distance of the atomic radius.

Evaluation of relative electronegativity was first accomplished by Pauling. He considered the energy of a heteronuclear single covalent bond to consist of the geometric mean of the homonuclear single bond energies of the two elements, *supplemented* by an electrostatic or ionic energy resulting from uneven electron sharing in the bond, which he attributed to an electronegativity difference between the two elements. He used the difference between the observed bond energy and the average of the homonuclear energies as a measure of the ionic energy. From such differences for various pairs of atoms he established a relative scale in which the most electronegative element, fluorine, has the value 4.0, and within the same period are O 3.5, N 3.0, C 2.5, B 2.0, Be 1.5, and Li 1.0.

Mulliken defined electronegativity as the average of the "valence state" ionization energy and electron affinity. Gordy suggested that electronegativity is a measure of the electrostatic potential at the surface of an atom, expressed as the effective nuclear charge, $Z_{eff}e$ divided by the radius. Allred and Rochow modified this concept by considering the electrostatic force, $Z_{eff}e^2/r^2$, as the measure of electronegativity. Electronegativities have also been estimated from work functions of metals, from force constants determined by infrared spectroscopy, from nuclear magnetic resonance spectroscopy, and by other methods. When adjusted to the same arbitrary scale, conventionally that established by Pauling, these methods give values in surprisingly good agreement, with only a few minor discrepancies that remain controversial. The principal application of *Pauling scale electronegativities*, which are those almost universally given in textbooks, has been qualitative prediction of bond polarity. That is, a given bond between atoms initially differing in electronegativity is polar with a partial negative charge on the initially more electronegative atom. The degree of polarity increases with increasing electronegativity difference.

Far more valuable applications of electronegativity have been made using a different scale based on the relative compactness of the electronic clouds of atoms. Electronegativities thus derived are approximately a linear function of the square root of the Pauling scale values, and in this sense in generally good agreement. They (see accompanying table) have been used for quantitative estimation of the partial charges on combined atoms, which in themselves permit correlation of a vast quantity of chemical data and interpretations of many chemical phenomena. The partial charges in turn have been applied to the quantitative calculation of bond energy. Furthermore, more recently, a simple quantitative relationship between homonuclear covalent bond energy and electronegativity has been demonstrated. Experimental homonuclear bond energy can be used to calculate electronegativity, or vice versa.

RELATIVE ELECTRONEGATIVITIES OF SOME ELEMENTS
(Relative Compactness Scale)[a]

H	3.55	K	0.42	Rb	0.36	Cs	0.28
Li	0.74	Ca	1.22	Sr	1.06	Ba	0.78
Be	2.39	Zn	3.00	Cd	2.59	Hg	2.93
B	2.93	Ga	3.28	In	2.84	Tl(I)	1.89
				Sn(II)	2.31		
C	3.79	Ge	3.59	Sn(IV)	3.09	Tl(III)	3.02
N	4.49	As	3.90	Sb	3.34	Pb(II)	2.38
O	5.21	Se	4.21	Te	3.59	Pb(IV)	3.08
F	5.75	Br	4.53	I	3.84	Bi	3.16
Na	0.70						
Mg	1.56	Sc	1.30	Y	1.05	La	0.88
Al	2.22	Ti	1.40	Zr	1.10	Hf	1.05
Si	2.84	V	1.60	Nb	1.36	Ta	1.21
P	3.43	Cr	1.88	Mo	1.62	W	1.39
S	4.12	Mn	2.07	Tc	1.80	Re	1.53
Cl	4.93	Fe	2.10	Ru	1.95	Os	1.67
		Co	2.10	Rh	2.10	Ir	1.78
		Ni	2.10	Pd	2.29	Pt	1.91
		Cu	2.60	Ag	2.57	Au	2.57

[a]Values for the transitional elements are tentative estimates only.

Space does not permit a detailed description of the concepts and methods mentioned here, but an example of the results obtainable may illustrate the principles involved. Silica, SiO_2, consists of silicon atoms initially of 2.84 electronegativity, tetrahedrally surrounded by oxygen atoms, initially of 5.21 electronegativity, each of which bridges two silicon atoms. The principle of electronegativity equalization states that when two or more atoms initially different in electronegativity combine, their electronegativities become equalized to the geometric mean. For SiO_2 the electronegativity of the compound is $(2.84 \times 5.21^2)^{1/3} = 4.26$. The equalization is brought about through the uneven sharing of the bonding electrons. The oxygen being initially more electronegative, attracts more than a half share of the bonding electrons. By spending more than half time more closely associated with the oxygen, the bonding electrons impart a partial negative charge on the oxygen, expanding the cloud through increased repulsions and decreasing the electronegativity. Simultaneously the silicon atoms shrink because of reduced repulsions, and increase in electronegativity because of reduced shielding between nucleus and bonding electrons. The decrease in oxygen electronegativity from 5.21 to 4.26 is 0.95. If oxygen had acquired an electron completely the electronegativity would have dropped by 4.75. The partial charge on oxygen is defined as the ratio $0.95/4.75 = -0.20$ (minus because the electronegativity decreased). The silicon is left with a partial positive charge of 0.40.

The silicon-oxygen bond, like all heteronuclear bonds, can be treated *as if* its energy were partly covalent and partly ionic. Instead of ionic energy supplementing the covalent energy, as suggested by Pauling, the ionic energy *substitutes* for a part of the covalent energy. The ionic weighting coefficient is half the charge difference: $(0.40 + 0.20)/2 = 0.30$. The covalent weighting coefficient, $1.00 - 0.30$, is 0.70. For the covalent energy one takes the geometric mean, 59.7, of the homonuclear bond energies of silicon (53.4) and oxygen (66.7, as calculated from the O_2 molecule), multiplies by 0.70, and corrects for bond length by the factor (covalent radius sum)/(observed bond length) = 1.90/1.61.

This calculation gives 49.1 kcal/mole of bonds. The ionic energy is simply the conversion factor (to kcal/mole) 332, times the weighting coefficient 0.30, divided by the bond length 1.61, or 61.8 kcal/mole of bonds. The sum, 110.9 kcal, is the Si—O bond energy. Atomization of SiO_2 requires the rupture of four SiO bonds per formula unit; $4 \times 110.9 = 443.6$ kcal/mole for the atomization energy of SiO_2(c). Subtraction of the atomization energies 108.9 for Si and 119.2 for two O gives -215.5 kcal/mole for the calculated standard heat of formation of SiO_2(c). The experimental value is -217.7.

Electronegativity thus permits a quantitative interpretation of the bonding in SiO_2 or any other compound for which appropriate data are known. The same principles allow a superior alternative to the "ionic" model of nonmolecular solids, and offer high hope of eventually elucidating the thermochemistry of mineral substances in general.

ELECTRON EMISSION. The liberation of electrons from an electrode into the surrounding space. Quantitatively, it is the rate at which electrons are emitted from an electrode.

ELECTRONEUTRALITY. If one describes the properties of electrolytic solutions in terms of ionic species, one has to take account of the fact that the concentrations of all species are not independent because the solution as a whole is neutral.

One generally uses the symbol z_i to denote the charge on an ion measured in units of the charge of a proton (for example, for Na^+, $z = 1$; for La^{3+}, $z = 3$; for PO_4^{3-}, $z = -3$); z is also called the *charge number* of the ion.

If n_i is the number of moles of the ionic species i, the condition of electrical neutrality is

$$\sum_i z_i n_i = 0 \tag{1}$$

Alternatively if one uses the subscript + to denote positively-charged ions or *cations* and − to denote negatively charged ions or *anions*, then one may write (1) in the form

$$\sum_+ z_+ n_+ = \sum_- z_- n_- \tag{2}$$

ELECTRON GAS. The term electron gas is used to denote a system of mobile electrons, as, for example, the electrons in a metal which are free to move. In the free electron theory of metals, these electrons move through the metal in the region of nearly uniform positive potential created by the ions of the crystal lattice. This theory when modified by the Pauli exclusion principle, serves to explain many properties of metals, especially the alkali metals. For metals with more complex electronic structure, and semiconductors, the band theory of solids gives a better picture.

ELECTRON GUN. An electrode structure which produces and may control, focus, and deflect an electron beam.

ELECTRONIC DATA PROCESSING. A few decades ago, when digital computers were introduced into the business and scientific fields for data processing, the term *electronic data processing* (EDP) was widely used. Today, use of the word *computer* in some form, such as *computer processing,* essentially has replaced EDP in the English business and technical language. See **Digital Computer**.

ELECTRONICS. An all-inclusive type of term embracing the study, design, and application of devices whose operation is dependent fully or partially upon the characteristics and behavior of electrons. Since electricity is a fundamental quantity, in nature comprising electrons and protons at rest or in motion, any phenomenon or device of an electrical nature would also be covered by the umbrella term, electronics.

The word *electron* was first used in a paper by George J. Stoney in the July 1891 issue of *The Scientific Transactions of the Royal Dublin Society*, entitled, "On the Cause of Double Lines and of Equidistant Satellites in Spectra of Gases." The word *electronics* (Ger. *Elektronik*) was first used to describe the branch of physics now generally called physical electronics. That usage dates back almost to the discovery of the electron in 1897, as witness the names of two early journals in the field, *Jahrbuch der Radioaktivität und Elektronik* (founded in 1904) and *Ion: A Journal of Electronics, Atomistics, Ionology, Radioactivity and Raumchemistry* (1908). In the currently prevalent technological context, the adjective *electronic* and the noun *electronics* (Ger. *technische Elektronik*) date back only to the 1920s.

The discovery of thermionic emission and the utilization of this effect in vacuum tubes set the stage for a new field of technology which, although basically electrical in character, nevertheless differed from the traditional spheres of electrical engineering and the science of electricity and magnetism. Invention of the triode by De Forest in 1907 and the development of new uses for vacuum tubes (or valves), with their particular attributes, notably amplification and oscillation, and with their unique solutions for problems, notably in the communications field, markedly differentiated these devices from prior run-of-the mill electrical and magnetic hardware. Designers of the early 1900s were able to develop practical and important electrical equipment without benefit of formal knowledge of electron theory. In Marconi's experiments of the late 1800s, using magnetic detectors for radio waves, and his dramatic accomplishment of spanning the Atlantic with radiotelegraphy in 1901, there was a definite pointing toward what is now considered modern electronics. This was given further emphasis with the first application of the three-electrode vacuum tube to radio in 1912. The Radio Corporation of America was formed in 1919 to pursue radiotelegraphy developments for the United States, which, at that time, enjoyed cable communications with only two nations in Europe—Great Britain and France. It was at about this time that the ever increasing and tight bond between electronics and communication was formed, which continues to the present and into the foreseeable future.

The first of the "modern" digital computers, the Mark I, was completed in 1944 for the U.S. Navy by Dr. Howard H. Aiken of Harvard University, assisted by International Business Machines Corporation. This was an electromechanical machine controlled by electromagnetic relays, two-position devices developed by Bell Telephone Laboratories as switches for the telephone industry. Other Bell Labs products used in the Mark I were punched paper tape equipment teleprinters and other components developed for communications systems. In today's parlance this machine might be called "electronic," but in its day it did not meet the requirement of incorporating vacuum tubes. However, the ENIAC (Electronic Numerical Integrator and Calculator), also built during the early 1940s by Dr. John Mauchly and Dr. J. Presper Eckert at the Moore School of Electrical Engineering at the University of Pennsylvania in Philadelphia did meet the test, and it was commonly referred to as the first true electronic computer. Instead of electromagnetic relays, some 18,000 flip-flop vacuum tubes originally designed for radar and television were used. And thus the firm and everlasting association between electronics and computing was established.

The close association between electronics and the vacuum tube persisted for many years, until the development of semiconductors. In the 1941 edition of *The Encyclopedia Americana*, the entry on *Electronics* was largely confined to thermionic emission, applications of diodes, grid-controlled thermionic tubes, and triodes as detectors and amplifiers. The definition of electronics given was, "the term applied to a wide variety of applications of the electron theory to engineering practice," which, in retrospect, was an excellent definition for its time.

Specialized publications for the field of electronics commenced in the 1940s. The American Institute of Electrical Engineers and the Institute of Radio Engineers combined in 1962 to form the Institute of Electrical and Electronics Engineers. The new organization was composed of 34 specialized groups, most of them pertaining to some aspect of electronics. In a way, this merger of two professional groups recognized somewhat belatedly the existence of an electronics technology, but also perpetuated some demarcation of electronics from other aspects of electrical technology.

In an excellent, short review of electronics, Pierce suggested that "electronics has come to mean all electrical devices for communication, information processing, and control." Pierce points out that this would include pre-vacuum-tube devices such as the electric telegraph, which replaced such light-wave communication as semaphore telegraphs, signal flags, and heliographs. It would include the telephone. And, thus, according to this definition, electronics was born nearly 150 years ago with the first electric telegraphs (*Science*, **195**, 1092–1095, 1977).

As of the late 1980s, the foregoing definition of electronics has expanded even further. Although all classifiers of industrial activity may not agree, the *electronics industry* is now considered to include the following activities and end products:

1. Semiconductors—discrete semiconductors, integrated circuits, and optoelectronic devices.
2. Circuit Components (electric and electronic)—capacitors, character displays, connectors, crystals, electron tubes, magnetic and microwave components, filters and networks, power supplies, printed circuits, relays, resistors, switches, keyboards, transducers, cable, and wire.
3. Data Processing Equipment—computers (all types and sizes), data storage devices, data terminals, input/output peripherals, automated office equipment (copiers, typewriters, word processors), and optical disk drives, among numerous others.
4. Communications Equipment—message switchers, modems, multiplexers, facsimile terminals, fiber-optic systems, radar equipment, radio, television, and telecommunications equipment.
5. Industrial Measurement/Control Equipment—process controls, including sensors, valve and motor controls, displays, etc.—discrete-piece *manufacturing controls*, including numerical controls, programmable controllers, robots, machine vision, and testing equipment, among many other categories.
6. Computer-Aided Design and Engineering Systems (CAD and CAE).
7. Consumer Equipment—radios, stereo systems, television receivers, time pieces, microwave ovens, telephones, games and toys, personal calculators and computers, among others.
8. Software—to support many of the foregoing activities and products.

Scores of entries in this encyclopedia relate to electronic phenomena and electronic products. Consult alphabetical index.

Additional Reading

Allison, J.: "Electronic Engineering Semiconductors and Devices," McGraw-Hill, New York, 1990.
Bartlett, E. R.: "Electronic Measurements and Testing," McGraw-Hill, New York, 1992.
Brophy, J. J.: "Basic Electronics for Scientists," McGraw-Hill, New York, 1990.
Coombs, C. F., Jr.: "The Printed Circuits Workbook Series," 3 vols., McGraw-Hill, New York, 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.
Frederickson, T. M.: "Introduction to Random Electronics," McGraw-Hill, New York, 1988.
Graf, R. F.: "The Encyclopedia of Electronic Circuits," 1 vol., McGraw-Hill, New York, 1988.
Graf, R. T.: "Electronic Databook, 4th Edition, McGraw-Hill, New York, 1988.
Hintz, K. J., and D. Tabak: "Microtrollers," McGraw-Hill, New York, 1992.
Inglis, A. F.: "Electronic Communications Handbook," McGraw-Hill, New York, 1988.
Sclater, N. J., and S. Gibillsco: "Encyclopedia of Electronics," McGraw-Hill, New York, 1990.

ELECTRONIC SOUND. See **Musical Sound.**

ELECTRON LENS. An electron moving in an inhomogeneous electric or magnetic field in general follows a curved trajectory. It may be shown that the trajectories in certain fields are such that all electrons which pass through a given point subsequently pass through, or close to, a second point, whose location is fixed by the field strengths, and by the position of the first point. The paths of electrons in such a system therefore bear a striking resemblance to the light rays which pass through a lens, and the set of electrodes or of conductors which establish the necessary fields is known as an electron lens. The focal length of the lens is defined in the same way as is the focal length of an optical lens; it may be varied by changing the field strengths. As an example of an electron lens, a slit or hole in a conducting sheet acts to converge or diverge electrons if the electric fields on the two sides of the sheet differ in strength.

ELECTRON MICROSCOPE. The concepts that eventually led to the development of electron microscopes came out of the discovery of the wave nature of the electron in 1924. The effective wavelength of the electrons varies with accelerating voltage and is less than 1Å: $\lambda = \sqrt{(150/V)}$ Å. This short wavelength makes possible far better resolution and higher magnification in the electron microscope as compared with the optical microscope.

The lenses used in electron microscopes act on the beams of electrons in much the same way that ordinary glass lenses act on beams of light. Most electron microscopes have electromagnetic lenses, although electrostatic lenses also can be used. Application of a uniform axial magnetic field causes the electrons to travel in a spiral path and return to the axis as shown schematically in Fig. 1. Except for the spiral part of the motion, this behavior is just like that of a simple glass lens, and the equations for optical lenses apply, as shown in Fig. 2:

$$\frac{1}{a} + \frac{1}{b} = \frac{1}{f}$$

$$\text{Magnification} = \frac{b}{a}$$

However, in the electromagnetic lens, the angle of deflection and the focal length depend upon the strength of the magnetic field. This can be controlled by adjusting the current in the coils so that the magnification of the lens can be continuously varied over a broad range. See Fig. 3.

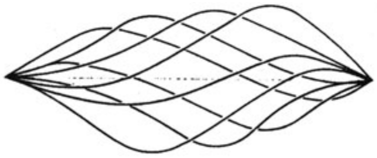


Fig. 1. Courses of electron beams in a homogeneous magnetic field.

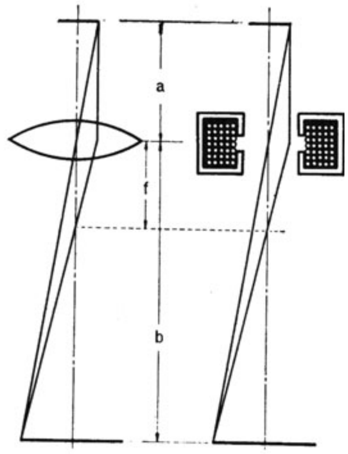


Fig. 2. Lens system used in transmission-type electron microscope. (Left) optical convex lens; (Right) electromagnetic lens.

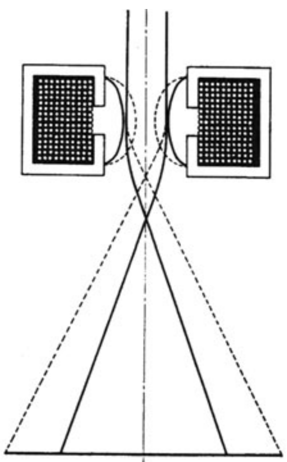


Fig. 3. Relation between magnetic field intensity and magnification in case of large magnetic field intensity (i.e., a large current flowing through the coil); or small magnetic field intensity (i.e., a small current flowing through the coil).

A simple electron microscope, as illustrated in Fig. 4, is operated by producing a beam of electrons from a heated filament, accelerating the beam with a high voltage applied to an anode, and then directing this beam of illuminating electrons onto the specimen. The specimen is in some respects similar to that used for optical microscopy, but because electrons are not so penetrating as light, the specimen must be much thinner (on the order of 1,000Å or less for most materials). For biological materials, these thin sections are produced by ultramicrotomes. Many materials, especially metals, can be thinned chemically or electrochemically. Rough surfaces can be examined by evaporating carbon on the sample and then removing it and using the carbon replica in the microscope.

The beam of electrons that passes through the specimen is then magnified by an objective lens and a projector lens and finally strikes either film held in a camera, or a fluorescent screen. The image seen on the fluorescent screen has varying shades of gray that depend upon the distribution of density and thickness of the specimen. This is most important for biological samples. For crystalline materials, the regular arrays of atoms aligned in critical directions can act like a mirror, diffracting the beam to another direction. This makes it possible to observe imperfections in the atomic arrangement of a material.

More complex electron microscopes use additional lenses, both above and below the specimen. The condenser lenses above the specimen concentrate the electron beam and increase the illumination. The addition of intermediate lenses below the specimen make it possible to go to higher magnification in the final image. Various alignment controls, apertures for the lenses, specimen handling devices, and suitable airlocks and anticontamination traps also are provided.

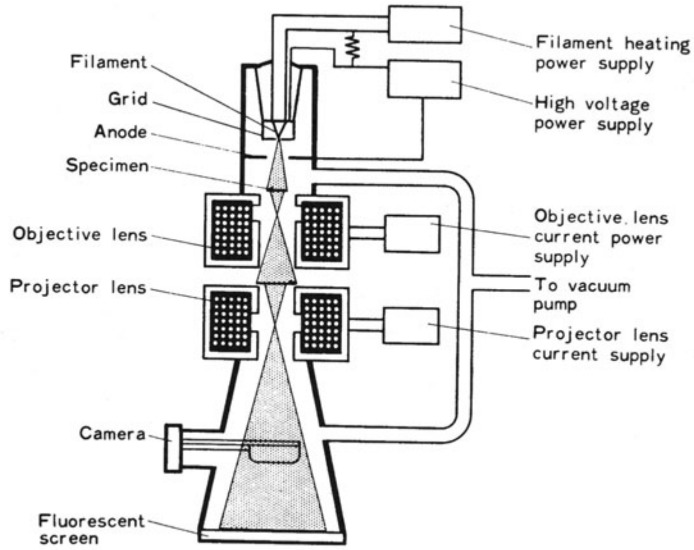


Fig. 4. Construction of a simple electron microscope.

The central column of the instrument must be maintained as a vacuum because the electrons would be absorbed if any atmospheric gases were present. Typical resolutions obtainable by commercial instruments are on the order of 2 to 5Å. Accelerating voltages from 20,000 to 1 million volts have been used. The higher accelerating voltages are useful for penetrating thick specimens (in some cases, up to 1 micrometer or more).

Scanning-Type Electron Microscope. This type of electron microscope is completely different in principle and application from the conventional transmission-type electron microscope. In the scanning instrument, the surface of a solid sample is bombarded with a fine probe of electrons, generally less than 100Å in diameter. The sample emits secondary electrons that are generated by the action of the primary beam. These secondary electrons are collected and amplified by the instrument. Since the beam strikes only one point on the sample at a time, the beam must be scanned over the sample surface in a raster pattern to generate a picture of the surface sample. The picture is displayed on a cathode ray tube from which it can be photographed.

A block diagram of a scanning-type electron microscope is given in Fig. 5. Major elements of the instrument include the electromag-

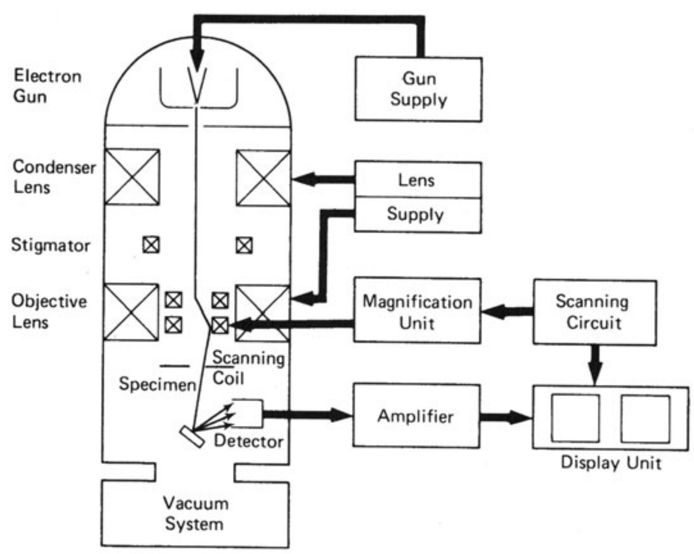


Fig. 5. Simplified diagram of scanning-type electron microscope.

netic lenses that are used to form the electron probe, the scan coils that sweep the beam over the sample, the detector that collects the secondary electrons, and the amplifying means where the secondary electrons are amplified and fed to the cathode ray tube for display. Since the cathode ray tube is scanned in synchronization with the electron beam, the resulting picture corresponds to the area of the sample being examined.

The advantages of the scanning microscope as compared with conventional optical microscopes include superior resolution and depth of field. Resolution of 200Å is obtained with a depth of field several hundred times that of a conventional optical microscope. The scanning microscope also makes possible the display of other kinds of data obtained from the specimen, notably cathodoluminescent photons emitted by fluorescing samples, electrical voltages generated in semiconducting samples by the passage of the electron probe, and characteristic x-rays that can be used to determine sample composition. Also see **X-Ray Analysis.**

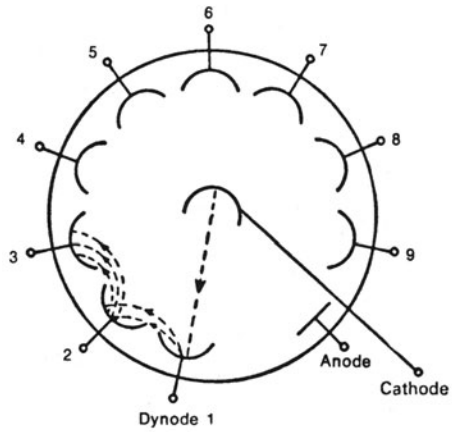
The range of magnification of scanning electron microscopes generally is from less than 30 × to about 40,000 ×, limited by the resolution of the instrument. Most specimens can be examined without any special preparation, but nonconducting specimens usually are coated with 100Å to 500Å of metal to conduct away the beam current. A conventional diffusion pump vacuum system is used since a vacuum is required for operation of the electron beam. The image that is formed is easily interpreted as surface topography inasmuch as the illuminating and shadowing effects on the sample are similar to the appearance of large objects as they normally are seen by the unaided eye. The scanning electron microscope has found broad application in the transistor industry to show voltage distributions in such devices. Other uses include industrial quality control and a broad range of industrial and biological research.

High-Resolution Scanning Transmission Electron Microscope (STEM). The concept for constructing a STEM dates back to 1963, growing out of the techniques and practices of research in nuclear and high-energy physics, with minimal consideration given to the large record of experience gained with the CTEM. The operating principles of the STEM reflect those of accelerator physics. As described by A. V. Crewe, a leading authority in the field, electrons from a field emission source are accelerated to a final potential V_0 and then focused on the specimen. Scattered electrons leaving the specimen normally are refocused by the magnetic field of the lens at some point farther down the column and then diverge. The elastically scattered electrons strike an annular detector; the inelastic and unscattered electrons are separated by a spectrometer. The beam is scanned across the specimen by using deflection coils; below the specimen, the scanning action is removed with additional deflection coils. The maximum scattering angle for the electrons is only 2° or 3°. Present resolution of the STEM is from 2 to 2.5 angstroms, still insufficient to resolve distances between atoms in most solids. A point resolution of 0.5 angstrom is needed to obtain images that will resolve such distances. See also **Scanning Tunneling Microscope.**

Additional Reading

Crewe, A. V.: "High-Resolution Scanning Transmission Electron Microscopy," *Science*, **221**, 325–350 (1983).
Gabriel, B. L.: "Biological Scanning Electron Microscopy," Van Nostrand Reinhold, New York, 1982.
Hawkes, P. W., Ed: "The Beginnings of Electron Microscopy," Academic Press, Orlando, Florida, 1985.
Olins, D. E., et al.: "Electron Microscope Tomography: Transcription in Three Dimensions," *Science*, **220**, 498–500 (1983).
Ottensmeyer, F. P.: "Scattered Electrons in Microscopy and Microanalysis," *Science*, **215**, 461–466 (1982).
Smith, D. J., et al.: "Dynamic Atomic-Level Rearrangements in Small Gold Particles (High-Resolution Electron Microscope)," *Science*, **233**, 872–875 (1986).

ELECTRON MULTIPLICATION. On bombarding certain surfaces by electrons it may happen that each impinging electron expels several electrons from the struck surface. If these electrons are caught in an electric field and driven against another similar surface, each of them may again give rise to several electrons. After several such stages of multiplication, an appreciable pulse may be obtained in this manner, starting with a single electron as beta radiation, or produced photoelectrically by gamma radiation. Quantitatively, the electron multiplication of a device such as a photomultiplier is the ratio of the number of electrons reaching the anode to the number emitted at the cathode.



Electron multiplier.

An electron multiplier is a device for amplifying by a process of electron multiplication. In the accompanying figure, electrons are emitted from the cathode and accelerated to an electrode called a dynode, maintained at about 100 volts positive with respect to the cathode. Secondary electrons knocked out of the dynode are accelerated to a second dynode at a higher potential, and there produce still more secondaries. The cathode, nine dynodes and final anode used in typical multiplier cells are so placed and shaped that the electron beams are efficiently focused at each stage. In typical multiplier cells, the gain per dynode is between four and five. The total gain is $(4)^9$ to $(5)^9$ or about 10^6. In addition to their use in photocells, multipliers are used in radiation detectors.

ELECTRON (Photoelectron) SPECTROSCOPY. A valuable research tool for conducting basic research in molecular chemistry and surface physics, chemical characterization of atmospheric particulate matter in environmental studies, and the analysis of the performance of catalysts in industrial processes, electron spectroscopy is based on the high-resolution analysis of the kinetic energy distributions of electrons emitted from solid, liquid, or gaseous substances upon irradiation with a beam of monoenergetic x-rays or ultraviolet radiation. Monochromatized synchrotron radiation has also become important as a tunable photo source. The physical quantity measured is the electron binding energy B, which is given by Einstein's relation $h\nu = B + K$, where $h\nu$ is the known photon energy and K is the measured photoelectron kinetic energy. Chemical information is obtained from chemically induced changes in the binding energies. In principle, all electron orbitals from the K shell out to the valence levels can be studied. When x-ray excitation is used, the technique is called x-ray photoelectron spectroscopy (XPS); when ultraviolet light is used, the term ultraviolet photoelectron spectroscopy (UPS) is used. The photoelectric effect was discovered by Hertz in 1883. K. M. Siegbahn, University of Uppsala, was awarded half of the 1981 Nobel Prize in Physics for his work in this field.

ELECTRON THEORY. In the 1830s, Faraday had tentatively suggested that his experiments in electrochemistry could be interpreted in terms of a small unit of charge attached to ions. This notion of individual "atoms of charge" was somewhat eclipsed, however, by the enormous success of Maxwell's theory of electromagnetism, which was generally interpreted, by 1880, as favoring a view that electrical phenomena were due to continuous charge distributions and motions. G. Johnstone Stoney, in 1874, and Helmholtz, in 1881, had suggested again an atomic interpretation of electricity, but it was not until the brilliant experiments of Perrin, J. J. Thomson, Zeeman, and others in the 1890s that the concept of the electron received firm experimental foundation. Later experiments and theory (Millikan, Bohr, etc.) established the constancy of the electronic charge and interwove the concept of an electron of definite charge and mass into the basic structure of the atom.

The Cathode Ray Controversy. After the discovery of the cathode ray in high-vacuum discharge tubes by Plücker in 1858, there developed, with the experiments of Goldstein, Crookes, Hertz, Lenard, and Schuster, a controversy over the nature of the rays. A predominately German school held that the rays were a peculiar form of electromagnetic rays. The British physicists thought they were negatively charged particles. The controversy provides a classic "case history" of the typical scientific controversy in which two quite different models both explain most, but not all, of the observable facts. For resolution of this controversy, see **Cathode Ray.**

Thomson's Determination of *e/m*. In 1897 Thomson devised an apparatus in which he could deflect a beam of cathode rays with a magnetic field of induction B and also with an electric field of strength E. If the fields are perpendicular to each other, and to the original path of the beam, and if they occupy the same region, then (with proper polarities and magnitudes of fields) the electric force on the beam can equal the magnetic force, so that the beam hits the same point on a fluorescent screen as when no fields are applied. If e is the charge of a given particle, m its mass, and v its velocity, $v = E/B$. Thus, velocities of typical cathode ray beams could be measured. If the magnetic field is used alone, and the curvature R of the beam is measured, then one can equate centripetal and magnetic field forces $mv^2/R = Bev$, and then deduce $e/m = v/BR$. With v known from the previous experiment, e/m can be calculated. Thomson's early values were not very precise, but later experiments of a similar type gave values close to 1.76×10^{11} coulombs/kg. More recent experiments, drawing on measurements of many kinds, give $e/m = (1.75890 \pm 0.00002) \times 10^{11}$ coulombs/kg.

The Zeeman Effect. In 1896 Zeeman discovered the broadening of spectral lines when a light source was in a strong magnetic field. Experimental refinements by Zeeman and others, and theoretical work by Lorentz and Zeeman, permitted the interpretation of this effect as due to the influence of the magnetic field on oscillating or orbiting negatively charged particles within the light-emitting or absorbing atoms. From the spectroscopic data, the ratio of charge to mass of these hypothetical particles could be shown to be equal to that of cathode rays. The Zeeman effect thus provided the first experimental evidence that the negative particles emitted by atoms when heated (Edison effect) or subject to high fields and/or ionic bombardment (cathode rays) or bombarded by short-wavelength light (photoelectric effect) were, indeed, actual constituents of the atoms and were probably responsible for the emission and absorption of light.

The Charge on the Electron. In the decade following 1897, many different methods were evolved for determination of ionic charges. Some methods depended upon measuring the total charge of a number of ions used as nuclei for cloud droplet formation. Other methods were more indirect—experiments, for example, which, combined with the kinetic theory of gases, could give crude values for Avogadro's number, N. By dividing the Faraday constant (the charge carried in electrolysis by ions formed from one gram-atom of a univalent element) by N, one could determine the average charge per ion. Similarly, the constants in Planck's theory of blackbody radiation, when evaluated experimentally, could provide a numerical value for N, as could certain experiments in radioactivity. All such methods gave values of N of the order of 6×10^{23}, and hence 1.6×10^{-19} coulomb for the ionic charge. None of these methods measured individual charges; strictly speaking, the value for the ionic charge could be thought of only as an average value.

Millikan's experiments with single oil drops, beginning in 1906, provided a method for measuring extremely small charges with precision. He was able to show that the charge on his drops was *always ne*, with $e = 1.60 \times 10^{-19}$ coulomb (modern value) and n a positive or negative integer.

He observed the motions of very small charged oil drops in uniform vertical electric fields. The drops were so small that they moved with constant velocity (except for Brownian fluctuations) for a given force. The force in each case was due to gravity acting on the mass of the drop and to the electric field (if any) acting on the charge, q, on the drop. The charge on a given drop could be changed by shining x-rays upon it. Using Stokes' law, in a form modified to correct for the fact that the drops were *not* large in comparison to the inhomogeneities of the surrounding air, and the velocity of a drop in free (gravitational) fall, Millikan could infer the diameter and mass of a given drop, and then calculate its charge. The charge q always equaled ne. A few other physicists, in similar experiments, thought they had detected electric charges smaller than Millikan's e, but their experimental techniques were probably faulty.

Millikan's experiment did not prove, of course, that the charge on the cathode ray, beta ray, photoelectric, or Zeeman particle was e. But if we call all such particles electrons, and assume that they have $e/m = 1.76 \times 10^{11}$ coulombs/kg, and $e = 1.60 \times 10^{-19}$ coulomb (and hence $m = 9.1 \times 10^{-31}$ kg), we find that they fit very well into Bohr's theory of the hydrogen atom and successive, more comprehensive atomic theories, into Richardson's equations for thermionic emission, into Fermi's theory of beta decay, and so on. In other words, a whole web of modern theory and experiment defines the electron. (The best current value of $e = (1.60206 \pm 0.00003) \times 10^{-19}$ coulomb.

The Wave Nature of the Electron. De Broglie had suggested in 1924 that electrons have in some respects the characteristics of waves, and deduced, for the wavelength equivalent to a moving electron, the expression $\lambda = h/mv$, in which m and v are the mass and speed of the electron and h is Planck's constant. If the electron is moving, for example, with a speed corresponding to 65 eV of energy, the corresponding "De Broglie wavelength" is 1.52 angstroms, which is in the x-ray range. This led Davisson and Germer to see whether electrons might be reflected from crystals after the manner of x-rays. They used a single crystal of nickel cut parallel to the (111) planes, and upon varying the electron speed at a fixed angle of incidence, they found not only a distinct "regular" reflection, but also a series of diffraction maxima strikingly similar to those obtained with the same crystal for x-rays of varying wavelength. The differences observed were satisfactorily explained as due to the refraction of the nickel for the electron waves.

G. P. Thomson independently reached the same conclusions in 1927. The hypothesis that matter exhibits both corpuscular and wavelike characteristics served as a stimulus for the formal development of quantum mechanics by E. Schrödinger, M. Born, W. Heisenberg, and others. Following the discovery, which eventually led to a Nobel Prize to Davisson and Thomson, electron diffraction was immediately utilized as a tool for the study of the structure of matter.

Other Characteristics of Electrons. In applying quantum mechanics to certain problems in atomic spectroscopy, in 1925 and 1926, Pauli, and Goudsmit and Uhlenbeck found that electrons must possess angular momentum of amount $\pm\frac{1}{2}(h/2\pi)$. Dirac's work on a generalized quantum theory of the electron showed that it possessed a related magnetic dipole moment of magnitude $eh/4\pi mc$. The ratio of the dipole moment to the angular momentum (e/mc) is larger than can be accounted for in classical terms with any homogeneous wholly negative model. The concept of electronic dipole magnetic moment is essential not only in spectroscopy but in theories of ferromagnetism (see **Magnetism**).

One may speak of the "classical radius of the electron," $a = e^2/mc^2$, derived by setting the self-energy of the coulomb field of a charge e contained at a radius a equal to the relativistic rest energy, mc^2 of the electron. This $a = 2.82 \times 10^{-13}$ cm, comfortably smaller than any atom, but larger than the usual estimates of sizes of protons and neutrons.

Positive Electrons. Dirac's paper in 1928 could be interpreted as predicting the existence of electrons that are positive. But until such particles were found experimentally by C. D. Anderson in 1932 in cloud chamber pictures of cosmic ray particle tracks, most physicists preferred other interpretations of Dirac's paper. Positive electrons, or positrons are now known (1) to occur as decay products from certain radioactive isotopes, (2) to be produced (paired with a negative electron) in certain interactions of high-energy gamma rays with intense electric fields near nuclei, and (3) to be the product of certain decays of certain mesons. In principle, positrons could form anti-atoms with nuclei made from anti-protons and anti-neutrons, but in practice almost all positrons produced in the observable universe quickly meet their end by annihilating themselves together with some hapless negative electron. The end product of a positron-electron annihilation is a pair of gamma rays.

ELECTRON TUBE. A device in which electrons are freed from the restraints of a solid conductor, pass across a free space (vacuum or gas at low pressure) and are again collected by a solid conductor, but during this passage in free space are controllable in manners which would be impossible if they had not been temporarily freed. Also known as valves (British), electron tubes were, until the perfection of semiconductor devices in the late 1940s and 1950s, the major components of nearly all electronic circuits and equipment. Although electron tubes continue to be used for certain applications, particularly involving specially-designed tubes, a massive replacement or substitution of transistors and other solid-state devices and approaches for electron tubes has taken place. Whereas early in the period of conversion from tubes to transistors, one would assume that a piece of electronic equipment incorporated electron tube circuits unless otherwise denoted, the assumption now is that electronic equipment circuitry will be solid-state unless otherwise specified.

The remarkable circuit and packing densities now obtainable in microelectronic devices have further deemphasized the electron tube. Nevertheless the technology built around electron tubes is classical and merits the brief condensation presented here.

The "revolution" in electronics required a number of years to achieve, recalling that the point-contact transistor was invented by J. Bardeen and W. H. Brattain (AT&T Bell Laboratories) as early as 1948. The earlier semiconductors were quite costly to produce and their mass production under very closely-controlled conditions was difficult to achieve. Good yield and quality control continue to be major problems among semiconductor manufacturers.

An electron tube consists of a heater for kinetic energy excitation of electrons, a cathode which acts as a transfer electrode source of electrons, controlling grid electrodes, and an anode that is maintained electrically positive with respect to the cathode. These elements are insulated from each other and enclosed within an evacuated envelope made of either glass, metal, ceramic, or a combination of these materials. A getter is flashed within the tube to absorb any residual gas molecules which could have a harmful effect, electrically and chemically, on the operation of the tube.

When the device has only two electrodes (a cathode and an anode), it is called a *diode*. With the anode maintained electrically positive with respect to the cathode, an electric field results which causes the electrons to move toward the anode. In the external circuit, the electrons flow from the anode through the load impedance and then through the voltage source to the cathode, which acts as a low-work-function transfer medium, and so back to the anode. The work function can be considered to be the total amount of work necessary to free an electron from a solid.

Other electrodes are introduced in some designs between the cathode and the anode in the form of grids. By varying the voltages on these intervening electrodes, it is possible to modify the electric field between the cathode and the anode, and thus to control the current in the external circuit. Tubes having one grid in addition to the cathode and anode are called *triodes*. Tubes with two grids are called *tetrodes*; and tubes with three grids are called *pentodes*. Generally, tubes are labeled in accordance with the total number of active electrodes in a linear arrangement using a common electron stream. Sometimes, two or more sections are enclosed within the same envelope (e.g., a diode-triode or a triode-pentode); these tubes are not referred to in terms of the total multielectrode structure, but they are designated in terms of the respective tube units.

For those readers who may be interested in much more detail pertaining to the once commonly used vacuum tubes, reference to the prior (6th Edition) of this encyclopedia is suggested.

ELECTRON VOLT. This is a convenient unit of energy for calculations in electronics and in connection with ionization or excitation of atoms or molecules. When an electric charge e is transferred from a region where the electric potential is V_1 to one where the potential is V_2, its potential energy changes by an amount equal to $e(V_1 - V_2)$. If the charge e is the electronic charge 1.602×10^{-19} coulomb (as it is when the transferred particle is an electron or a proton), and if the potential difference $V_1 - V_2$ is one volt, the corresponding change in energy is equal to 1.602×10^{-19} joule or 1.602×10^{-12} erg, and is called an electron volt. Thus if a doubly ionized positive oxygen molecule moves in an electric field through a potential drop of 500 volts, it receives $2 \times 500 = 1000$ electron volts or 1.602×10^{-9} erg of energy; and since the mass of the oxygen molecule is about 5.31 $\times 10^{-23}$ grams, this energy would give the molecule, if unimpeded, a speed of about 7.77×10^6 centimeters per second or 48.1 miles per second. The abbreviation for electron volt is eV. In x-rays, nuclear physics, and elementary particle physics, the use of higher energies is encountered and additional abbreviations in common use are keV for 10^3 eV, MeV for 10^6 eV, and GeV (sometimes BeV in the United States) for 10^9 eV.

ELECTROOSMOSIS. The movement of liquid with respect to a fixed solid (e.g., a porous diaphragm or a capillary tube) as a result of an applied electric field. See also **Drainage Systems; Electroendosmosis;** and **Solar Energy.**

ELECTROPHILIC REACTION. The reaction in which an electrophilic reagent attacks a nucleophilic compound. The reagent is taken to be the inorganic substance (in the case of reactions of inorganic and organic substances) or the simpler of two reacting organic compounds. The electron-pair for the bond formed is furnished by the nucleophilic compound. The term *electrophilic* connotes *electron-seeking* and is applied, for example to positively-charged cations, or to reactions brought about by them.

ELECTROPHORESIS. The movement of a charged particle in response to an electric field when placed in that electric field, also sometimes referred to as cataphoresis or kataphoresis. Although electrophoresis in concept is several decades old, it remains as a dominant instrumental approach in studies of proteins, DNA, and other biological substances. The core techniques for separating large and small fragments of DNA, for sequencing DNA, and for the analysis of complex mixtures of proteins involve electrophoresis in microporous gels. As pointed out in the entry on **Gene Science,** it is now technically feasible to map and sequence the entire human genome and to resolve and identify all of the protein gene products. A major effort in this direction is mounting in the United States—with parallel efforts in Japan, England, Germany, and France. As pointed out by Anderson (1987), the most useful index to DNA databases will be based upon gene products, inasmuch as genes are usually named for the proteins they code for, once they have been described. Thus, high-resolution, two-dimensional (2D) protein electrophoresis and the data associated with 2D methods will be strongly linked to DNA databases. To discern part of the coding DNA sequence, the key experimental technique in use by investigators is the electrophoretic transfer of 2D protein patterns to activated glass, after which the transferred spots are partially sequenced. The method allows the investigator to search the DNA database for the gene coding for a particular spot, and the gene identified may be isolated and cloned. Given a complete human sequence library, the gene for every protein resolved by 2D electrophoresis can be found.

Anderson continues in his observation that the data storage requirements for 3.4×10^9 base pairs of DNA and 30,000 to 100,000 proteins and the databases and computational and display systems required are not considered large compared with those now in use in physics, by the military, and by the intelligence community. New, large-scale robotic DNA sequencing and protein-mapping systems, however, will be required. Except in large centralized laboratories for clinical chemistry, large-scale bioanalytical systems have not been systematically developed.

Principles of Electrophoresis. It was observed, many decades ago, that when electricity is passed through a solution containing colloidally dispersed particles, the negatively charged particles move toward the positive electrode, and positively charged particles move in the opposite direction. Moreover, the particles move at differing speeds, depending on such properties as their net electric charge, size and shape, thus making it possible to separate them from a mixture. The charge on a particle may arise from charged atoms or groups of atoms that are part of its structure, from ions which are adsorbed from the liquid medium, and from other causes. It soon became evident that the behavior of colloidal particles in an electric field, as compared to that of ions, differed in degree rather than in kind. Although a colloidal particle is much larger than an ion, it may also bear a much greater electrical charge with the result that the velocity in an electric field may be about the same, varying roughly from $0–20 \times 10^{-4}$ centimeter/second in a potential gradient of 1 volt/centimeter.

To understand the phenomenon of electrophoresis, let us suppose, for simplicity, that a non-conducting particle, spherical in shape, of radius r, and bearing a net charge of Q coulombs, is immersed in a conducting fluid of dielectric constant D, having a viscosity of η poises. Suppose, further, that the particle moves with a velocity of v centimeters/second under the influence of an electrical field having a potential gradient of x volts/centimeter. The force causing the particle to move, namely $Qx \times 10^7$ dynes, is opposed by the frictional resistance offered to its movement by the liquid medium. From Stoke's law, the latter is given by $6\pi\eta rv$. Under steady-state conditions, and by introducing the electrophoretic mobility $u = v/x$, rearrangement yields the expression $u = Q \times 10^7/6\pi\eta r$. It is evident that if the electrophoretic mobility of a particle can be computed, it should be possible to determine Q, the net charge on the particle.

For the micro and moving-boundary techniques, a more rigorous treatment of the problem must take into account such complicating factors as electroosmosis, the actual size and shape of the moving particle, the electrolyte concentration in the solvent medium, and the conductivity of the particle itself. Although the ionographic technique is simple from an experimental standpoint, additional complex factors are introduced due to the presence of the stabilizing agent, namely paper, cellulose acetate, starch, etc. However, it is now possible to introduce suitable corrections for these factors and to arrive at mobility data of sufficient quality to be useful in physical chemical computations.

Early Methodologies. Prior to the current emphasis on gene science, electrophoresis had become established as an important means for analyzing naturally occurring mixtures of colloids, such as proteins, lipoproteins, polysaccharides, nucleic acid, carbohydrates, enzymes, hormones, and vitamins. Also prior to chromatography, electrophoresis offered the only available method for the quantitative analysis and recovery of physiologically active substances in a relatively pure state. Electrophoresis offered a convenient and dependable means for analyzing protein content of body fluids and tissues and thus was (and continues to be) used widely in clinical and hospital laboratories. For example, the marked differences between normal and pathological serum samples are useful in the diagnosis and better understanding of disease. Such changes in the electrophoretic pattern of blood serum are evident in diseases characterized by marked protein abnormalities, such as multiple myeloma, nephrosis, obstructive jaundice, liver cirrhosis, and various parasitic disorders. Because of the small amount of fluid required, the method is applicable to the study of spinal fluids.

Prior to the 1950s, several methods were used for studying the electrophoretic behavior of charged particles in a liquid. Some of these methods are now essentially obsolete.

Microscopic Method. In this method, the migration of particles was observed in a solution contained in a glass tube placed horizontally on the stage of a microscope. The method was suitable for the study of relatively large particles, such as bacteria, blood cells, or droplets of oil. The method was markedly extended by coating the substances of interest onto tiny spheres of glass, quartz, or plastic. When completely coated, these spheres acted as if they were large protein particles and responded to an electrical field in terms of the charge on the protein. The method is now limited to historical interest.

Moving-Boundary Technique. In this method, the movement of a *mass* of particles is measured, thus obviating the need to observe individual particles. The displacement of the particles in an electric field is recorded photographically as the movement of a boundary between a solution of a colloidal electrolyte, such as a protein, and the buffer against which it was dialyzed. Material to be studied is poured into the bottom of a U-tube, and on top of it; in each arm of the U-tube, a buffer solution is carefully layered so as to produce sharp boundaries between the two solutions. Electrodes, inserted in the top of each arm of the tube, are attached to a direct current source.

The method was improved by Tiselius and co-workers (Sweden). Numerous other advancements have led to what might be termed contemporary electrophoretic methods.

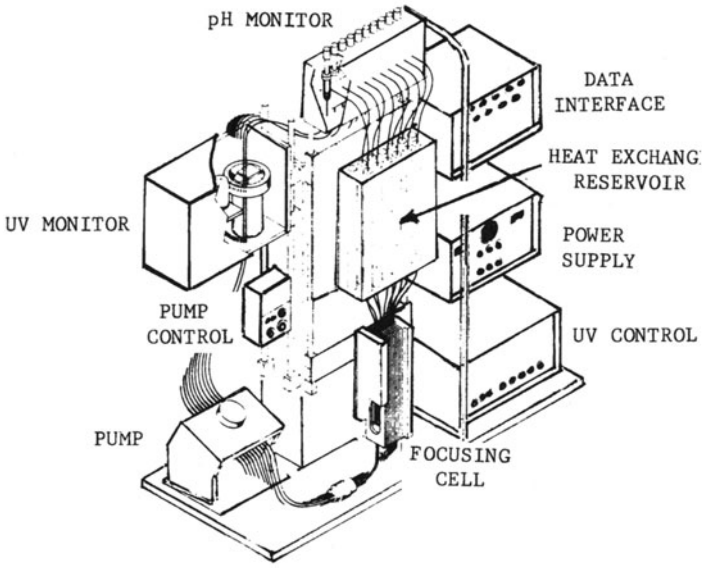
Contemporary Methods. Rigid restrictions on temperature, current, and composition of solutions are largely removed when electrophoresis is carried out in a solution stabilized with a material such as paper, cellulose acetate, starch, polyacrylamide gel, or agar. In addition, only minute amounts of material are required. There has been considerable specialization in the design of electrophoresis equipment in recent years. *Electrofocusing*, for example, is a very-high-resolution technique which separates biomolecules on the basis of their intrinsic charge. Used chiefly for peptides and proteins, electrofocusing is fast (one-half to one-third of the time required by traditional electrophoretic separations) and it has a high sample capacity capable of separating nearly 100 samples on one gel. Analytical electrofocusing is widely used for quality control. Systems have been designed to employ pH gradients—which are immobilized into the gel matrix during gel polymerization and thereby eliminating gradient drift and providing resolution of biomolecules differing by as little as 0.001 pH units in their isoelectric points. Another special-purpose system is based upon highly specific and sensitive antigen-antibody reactions. Known as immunoelectrophoresis, this allows both the qualitative and quantitative analysis of even trace components in a highly complex mixture. With the growing importance of monoclonal antibodies, immunoelectrophoresis has received wide acceptance. *Isotachophoresis* is a specialized form of electrophoresis in which the separation is achieved based on differences in electrophoretic mobility. This represents a high-resolution method for quantitative/quality control of ions including proteins, peptides, nucleosides, nucleotides, and other types of metabolites. Isotachophoresis also provides an effective method for environmental monitoring of hazardous agents found in the biotechnology industrial workplace, especially mutagenic or carcinogenic potentials, such as aromatic hydrocarbons. Isotachophoresis offers several advantages over alternative electrophoretic and chromatographic procedures, including short analysis time and high sensitivity.

Two-dimensional electrophoresis, introduced in 1976, combines polyacrylamide (sodium dodecyl sulfate-polyacrylamide) gel electrophoresis with isoelectric (electrofocusing) electrophoresis. The method is attractive because it can separate large numbers of proteins. A more recent two-dimensional technique is called *electrophoretic titration curve analysis*. As reported by Maugh (1983) a pH gradient from 3 to 9 is established in a horizontal slab. A trough is then cut into the slab at a right angle to the gradient. The sample is applied to the trough and conventional electrophoresis is performed, perpendicular to the gradient to produce a classical titration curve. The latter provides useful information about the protein, including its stability and the binding of ligands. The data also predict behavior of the protein in ion-exchange chromatography.

Another recent technique, that of preparative electrophoresis, is *recycling isoelectric focusing* developed by Bier and colleagues (University of Arizona). See accompanying diagram.

A general trend in electrophoresis is to use thinner gels to provide better retention of nucleotides for DNA sequencing and, because of more efficient cooling, higher voltages can be used for faster separations. Prior gels typically have been 2 mm in thickness; current gels are typically less than 0.3 mm thick. Numerous equipment and procedural advancements are continuing apace.

Capillary Electrophoresis. This technique, which appeared in the late 1980s, makes possible rapid and automated analysis of small volumes of complex mixtures with excellent resolution and sensitivity. The procedure is well described by M. J. Gordon et al. (see reference listed).



Schematic representation of the recycling isoelectric focusing (RIEF) apparus developed by Milan Bier (University of Arizona). In this preparative electrophoresis technique, no gel is used; instead the solution is recycled to be fractionated through a focusing cell and heat-exchange reservoir. The focusing cell is a series of parallel-flow chambers which are separated by monofilament nylon screen elements. The screens streamline the flow through the apparatus and avoid loss due to convection. The proteins are allowed to migrate from chamber to chamber exclusively under the influence of the electric field. Each chamber is connected to a separate glass channel in the heat-exchange reservoir. Capacity of the apparatus is determined by the volume of the heat exchanger and the cross-sectional area of the focusing chambers. A modular design permits scale-up to larger volumes. The pH gradient is established by the electric field itself, in the same manner as in analytical gel focusing. The apparatus has been used to process milligram quantities of antibodies to single-band purity within a period of 4 hours. (*M. Bier.*)

Additional Reading

Deutsch, J. M.: "Theoretical Studies of DNA During Gel Electrophoresis," *Science*, 922 (May 13, 1988).

Dose, E. V., and G. A. Guiochon: "Internal Standardization Techniques for Capillary Zone Electrophoresis," *Analytical Chemistry*, 1154 (June 1991).

Gordon, M. J. et al.: "Capillary Electrophoresis," *Science*, 224 (October 14, 1988).

Jorgenson, J.: "Recent Trends in Electrophoresis," *Science*, Part II, page G82 (February 12, 1988).

Kennedy, R. T. et al.: "Microcolumn Separations and the Analysis of Single Cells," *Science*, 57 (October 6, 1989).

Smisek, D. L., and D. A. Hoagland: "Electrophoresis of Flexible Macromolecules," *Science*, 1221 (June 8, 1990).

Smith, S. B., Aldridge, P. K., and J. B. Callis: "Observation of Individual DNA Molecules Undergoing Gel Electrophoresis," *Science*, 203 (January 13, 1989).

ELECTROPHORUS. The simplest of all static machines; devised by Volta in 1816. It consists of a slab of some resinous substance, such as sealing wax or vulcanite, which is negatively charged by rubbing with fur. A metal plate provided with an insulating handle is placed upon the electrified slab. The contact is localized at a few points, so that instead of taking the negative charge off the slab, the metal plate becomes charged by induction, positively on the under side and negatively on the upper. The negative induced charge is now removed by grounding with the finger, and upon being lifted by means of the handle, the plate becomes positively charged all over, often strongly enough to yield bright sparks. Very little of the negative charge on the slab is removed in this process, and it may thus be used over and over to induce an indefinite number of positive charges. The energy is of course furnished by the operator in pulling the metal plate away from the slab. If a slab of glass is used, and rubbed with silk, it becomes positive and the induced charges on the plate are then negative. The instrument is useful in lecture-table demonstrations.

ELECTROPLATING. The coating of an object with a thin layer of some metal through electrolytic deposition. The process is widely used, either for the purpose of rendering a lustrous or noncorrosive finish on some article. In electroplating, the general object is to employ the article to be plated as the cathode in an electrolytic bath composed of a solution of the salt of the metal being plated. The other terminal, the anode, may be made of the same metal, or it may be some chemically unaffected conductor. A low-voltage current is passed through the solution, which electrolyzes and plates the cathodic articles with the metal to the desired thickness. In this way, table utensils are silver plated, various parts are made weatherproof by cadmium or chromium plating, and a high finish may be imparted through nickel plating. Copper, zinc, and gold are also plated. As the plating proceeds, the strength of the solution must be kept up by the addition of crystals of the plating salt, or a renewal of the anode if it is of the plating metal. A firm bond between the anode and the deposited metal is to be secured when the two metals are of a type which tends to alloy. If this is not the case, some intermediate metal, which will alloy between the base and the plate, is first deposited. For example, in silver plating on steel, the iron would otherwise form a poor bond with the silver, so a thin layer of copper is first deposited on it.

Because of the excellent conducting properties of a metallic salt solution, only a low voltage is required. As this must be dc, the process of electroplating calls for a supply of current from special low-voltage dc generators or rectifiers. The voltage will be of the order of 6 volts or less between anode and cathode.

Some of the solutions used to place various metals are as follows: for silver or gold plating, double cyanide of the metal and potassium; copper plating, copper sulfate; nickel plating, nickel ammonium sulfate. The articles to be plated must be thoroughly and effectively cleaned of all grease and dirt by washing in caustic or acid solutions. While the above is a brief outline of the process of electroplating, in commercial operations there are many troublesome angles which would not be suspected from the foregoing. Irregularity of the plate, poor surface graining, "trees," insufficient bond, and other troubles develop. The overcoming of these requires the use of various expedients, such as careful control of temperature, or the addition of certain colloids and other compounds which have been found effective in preventing formation of defects on the plated articles. See accompanying table.

METALLIC IONS COMMONLY ELECTROPLATED

Ion	E^0	Ion	E^0
Zn^{+2}	−0.762	Cu^{+2}	+0.345
Cr^{+3}	−0.71	Cu^{+1}	+0.522
Cd^{+2}	−0.402	Ag^{+1}	+0.800
Ni^{+2}	−0.250	Au^{+3}	+1.42
Sn^{+2}	−0.136	Au^{+1}	+1.68
(H^{+1}	0.000)		

As will be noted from the table, in order to plate out the metals above hydrogen from an aqueous solution, the concentration of the hydrogen in the cathode film must be low. Hence a basic solution, such as provided by a cyanide bath, may be required. A cyanide bath also may be used to provide a smooth adherent plate.

To plate out an alloy, that is, to *codeposit* at least two kinds of metals, the plating bath must be so designed that the electrodeposition potentials of the two metals in the cathode film are equal or nearly so.

Four typical electroplating baths are:

Copper:	CuCN 26 g/l, NaCN 35 g/l, Na_2CO_3 30 g/l $KNaC_4H_4O_6 \cdot 4H_2O$ 45 g/l, NaOH to give pH of 12.6
Copper:	$CuSO_4 \cdot 5H_2O$ 188 g/l, H_2SO_4 74 g/l
Tin:	Tin (as tin fluoborate concentrate) 60 g/l, free fluoboric acid 100 g/l, free boric acid 15 g/l.
Zinc:	$Zn(CN)_2$ 60 g/l, NaCN 23 g/l, NaOH 53 g/l

The bright, hard, ornamental chrome so popular on automobiles and household and office articles is produced by electroplating. In this process, the chromium is not present in the bath as a positive metal ion, but rather as part of the anion of chromic acid, H_2CrO_4. The object being plated is made the cathode. Usually the article will have previously been plated with copper and then with nickel. Sulfuric acid serves as a catalyst. The final chromium plate ranges from 0.00001 to 0.0005 inch (0.025–0.127 mm) in thickness. Although improvements have been made in recent years for certain chromium-plating applications, the traditional bath is an aqueous solution of chromic trioxide, CrO_3, and sulfuric acid, with a ratio of approximately 100 to 1 (wt). Fluosilicate catalysts are also added to some chromium-plating baths. See also **Chromium.**

The voltage for electrodeposition under ideal conditions would be:

$$E = E^0 + \frac{RT}{VF} \ln A$$

where E is the required voltage relative to the solution as measured by a hydrogen electrode in a unit molal activity hydrogen ion solution. E^0 is the electrolytic potential, in volts, of the metal being plated when immersed in a solution containing its ions at unit molal activity (approximately unit molal concentration). $R = 8.31$ joules per degree mole; T is the Kelvin temperature; F 96,500 joules per gram-equivalent; V is the valence of the ions which are depositing out; ln A is the natural logarithm of the activity of these ions (approximately the natural logarithm of their molality).

In actual practice the concentration and therefore the activity of the ions soon after the electroplating process starts is different in the solution just next to the cathode, called the cathode "film," than in the main body of the bath. The foregoing equation must in practice be modified to read:

$$E = E^0 + \frac{RT}{VF} \ln A - P,$$

where A is the molal activity in the cathode film of the ions being electrodeposited and P is the extra potential required to keep the plating going. A and P depend on temperature, current, density, concentration, valence, pH, and ion mobility.

P. R. Albright

ELECTROPOLISHING. Production of a smooth surface on metals by electrochemical means.

In all electroplating processes, metals (and hydrogen) are deposited on the cathode and dissolved from the anode, except when insoluble anodes are used in which case oxygen is liberated at the anode. Electropolishing is the reverse of electroplating. The work is made the anode and tends to be dissolved. The operating conditions are controlled so that atomic oxygen forms continuously and reacts with the metal surface. Part of this oxygen may be liberated as gas. According to one theory, the high points of the metal surface are most readily oxidized and this oxidized material is thereupon dissolved in the electrolyte or otherwise removed. In any case, selective solution of the high points of the surface tends to give a very smooth finish comparable or superior to a mechanically buffed surface. A wide variety of electrolytes is used. A typical one for stainless steels contains phosphoric acid and butyl alcohol.

All mechanical methods of polishing, including those used for metallographic samples, produce a thin surface layer of work-hardened metal. Electropolishing produces a strain-free surface which is especially suitable for microscopic examination.

An important commercial application of the process is the polishing of stainless steel parts of irregular contour which would be difficult or impossible to buff. Copper and its alloys, Monel metal, aluminum, and many other alloys can be electropolished.

ELECTROSCOPE. An instrument for detecting small charges of electricity, or for measuring small voltages, or sometimes, indirectly, very small electric currents, by means of the mechanical forces exerted between electrically-charged bodies. One of the earliest sensitive electroscopes consists of two narrow strips of gold-leaf hanging together in a glass jar. Upon being charged, they stand apart on account of their mutual repulsion. One leaf may be replaced by a stiff strip of brass, so that only the remaining leaf can move. See also **Electrical Instruments;** and **Electrometer.**

ELECTROSOL. A colloidal solution produced by electrical means, as by passing a spark between metal electrodes in a liquid.

ELECTROSTATIC GENERATOR. Any apparatus that generates electrostatically a voltage between two terminals. If a sufficiently large electrostatic charge can be accumulated on a particle large enough to be seen visually, it can also be accelerated to measurable speeds by an electrostatic generator. Such acceleration of particles provides a means for laboratory studies of the effects of collisions in space between satellites and micrometeoroids. A particle of mass m (in kilograms) carrying a charge q (in coulombs) and accelerated through a potential difference of V volts attains a speed $v = (2Vg/m)^{1/2}$ meters/second. Carbonyl iron spheres one micrometer in diameter, and carrying a charge up to about 3×10^9 volts/meter, have been accelerated to speeds in the 5 to 6 km/sec range in a 2 million volt accelerator. Improvements in particle charging techniques are expected to bring on increases in particle speeds into the hypervelocity range attained by real micrometeoroids.

One type of electrostatic generator is known as a Van de Graaff® generator in which a moving belt is charged by a low-voltage supply, such as a battery, and this charge is deposited onto a hollow, spherically shaped shell, on which a voltage can be developed that is 100 times or more the voltage of the primary supply, depending on the radius of the shell and the electric field at which voltage breakdown occurs. Positive ions or electrons, depending on the sign of the charge on the shell, can then be accelerated from a source inside the shell down an evacuated tube to strike a target at the grounded end of the tube. These charged particles strike the grounded terminal with an energy eV equal to the voltage applied at the source.

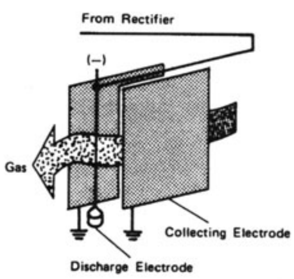
ELECTROSTATIC LENS. An arrangement of electrodes so disposed that the resulting electric field produces a focusing effect on a beam of charged particles.

ELECTROSTATIC PRECIPITATOR. The concept of the electrostatic precipitator for the removal of particulates from smoke and industrial emissions dates back to the late-1800s and the pioneering work of Frederick Gardner Cottrell. Some of the earliest work in connection with electrical precipitators was directed to copper smelting in the early 1900s. Quoting briefly from reports, "In the first decade of the century there were three well-publicized and classic examples of smelter smoke injury in the United States—at Ducktown, Tennessee, at Salt Lake City, and at Anaconda, Montana. Some indication of the scope of the problem can be seen from a study that was eventually made at the latter and which revealed some startling figures. In a normal day's operation, up and out the stack of that copper smelter went the amazing total of 3,200 metric tons of sulfur dioxide, 200 metric tons of sulfur trioxide, 30 metric tons of arsenic trioxide, 3 metric tons of zinc, and over 2 metric tons each of copper, lead, and antimony trioxide. The marvel is that anything remained, but nothing could more clearly demonstrate that here in full bloom was a cardinal essence of successful invention: a need existed. . . . The emissions from the low stacks of an old plant operated at a neighboring location had killed all vegetation, and losses of livestock by arsenical poisoning had been heavy over the near-lying area. Years after the plant was dismantled, the topsoil of a large area centering at the old site was stripped off, sent through concentrators, and smelted at the new plant with a reported recovery of over $1 million in copper and other metals."

But despite the need and the invention, Cottrell had considerable difficulty over a number of years in gaining acceptance of the electrical precipitator by industry. Today, and for a number of decades past, the electrostatic precipitator has been a major device for combatting air pollution. Since the precipitator functions only against particulates, numerous other items of air pollution control equipment, such as absorbers, scrubbers, and filters, are required and are described elsewhere in this volume.

Operating Principle. The basic operating principle of the electrostatic precipitator is demonstrated by the familiar experiment in which a glass rod is rubbed with a silk cloth; the action gives the rod an electrostatic charge, making it capable of attracting uncharged bits of paper, lint, or cork. In the electrostatic precipitator, it is the collecting surfaces that are grounded, while the charge is created on the particulates to be collected.

The power supply is a transformer-rectifier set which steps up ordinary 220-V ac supply to the high level necessary for precipitator operation, and rectifies it to direct current. The dc voltage is applied to discharge electrode wires suspended in the gas flow path. See accompanying figure. In the most common industrial type of precipitator, the discharge electrodes hang between rows of collecting electrode plates which form a series of parallel gas flow ducts. The high potential on the discharge electrodes causes a corona discharge, from which electrons migrate out into the gas. These create gas ions, which attach themselves to particulates in the gas and give the particles a charge.



Arrangement of electrodes in electrostatic precipitator.

The collecting electrodes are grounded, so that the high potential difference between them and the discharge electrodes creates a powerful electric field through which the gas must flow. According to Coulomb's law, such a field exerts a force on charged particles in the field. In the precipitator, this force moves particles out of the gas stream to the collecting electrodes. In a typical precipitator, the force on a particle 0.5 micrometer in diameter is several thousand times the force of gravity on such a particle.

At the grounded collecting electrodes, the particulates lose their charge. They drain off, if liquid, accumulate until washed off or, more commonly in the case of dry dust, are dislodged by mechanical agitation of the electrodes. In a few applications, the collecting electrodes are vertical pipes, instead of parallel plates, each pipe with a discharge electrode wire hanging down its axis.

ELECTROSTATICS. That branch of electromagnetism that deals with the effects of stationary (as opposed to moving) electric charges. The basic law of electrostatics is the Coulomb law. Materials are conveniently classed as conductors and nonconductors. In the former, charges are free to move within the conductor and any charge placed on a conductor will so distribute itself over the surface that the electric field within the conductor is zero and so that the conductor is an equipotential. When an uncharged conductor is placed in an electric field, produced, for example, by a neighboring charged body, a separation of charge on the surface will occur. A charge placed on a nonconductor will remain where it was placed. No material is a perfect nonconductor, but may approximate one very closely. When a nonconductor is placed in an electric field, electric dipoles are induced within it.

Static electric charges may be built up on a body by friction, by electrostatic induction, and by other means.

The laws of electrostatics are expressed in terms of electric charges, q, charge densities $\rho = dq/dV$, electric field strengths $\mathbf{E} = \mathbf{F}/q$, and electrostatic potentials $\phi = -\int \mathbf{E} \cdot ds$. The basic law of electrostatics is the Coulomb law of force between charges:

$$\mathbf{F} = \frac{q_1 q_2\ \mathbf{r}}{4\pi\epsilon_0 r^3},$$

with the resulting field due to a single point charge q:

$$\mathbf{E} = \frac{q\mathbf{e}_r}{4\pi\epsilon_0 r^2} \text{ (rationalized mksa units),}$$

where $\mathbf{e}_r$ is a unit vector pointing in the direction of increasing $\mathbf{r}$. The field due to a number of discrete charges is the vector sum

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0} \sum_i \frac{q_i(\mathbf{e}_r)_i}{r_i^2}$$

while that due to a continuous distribution of charges is

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0} \int \frac{\rho e_r dV}{r^2}.$$

The potential at some point in space due to a single point charge, referred to an origin of potential at infinity, is $\phi = q/4\pi\epsilon_0 r$, while the potentials respectively due to a number of discrete charges or to a continuous distribution of charges are

$$\phi = \frac{1}{4\pi\epsilon_0} \int \frac{\rho\, dV}{r} \quad \text{and} \quad \phi = \frac{1}{4\pi\epsilon_0} \sum_i \frac{q_i}{r^2},$$

again referred to an origin of potential at infinity. If the potential function ϕ has been determined, then the electric field may be determined from $\mathbf{E} = -\nabla\phi$ where $\nabla\phi$ is the gradient of ϕ. Alternatively, if the field $\mathbf{E}$ has been determined, the potential ϕ may be obtained from $\phi = \int_\infty^R \mathbf{R} \cdot d\mathbf{s}$.

The basic law of electrostatics, the Coulomb law, may alternatively be expressed either as the Gauss law or as the Poisson equation. In the absence of local charges, the Poisson equation reduces to the special case of the Laplace equation.

The electrostatic field is a conservative field, which leads to the fact that is possible to set up a potential function ϕ. This fact may be alternatively stated in terms of the closed line integral $\oint \mathbf{E} \cdot d\mathbf{s} = 0$, or in terms of the curl of $\mathbf{E}$: $\nabla \times \mathbf{E} = 0$.

ELECTROVALENCE. See **Chemical Elements.**

ELECTROSTRICTION. The variation of the dimensions of a dielectric under the infuence of an electric field.

ELECTROVISCOUS EFFECT. The change in viscosity of a liquid when placed in a strong electrostatic field. The effect is very small and occurs only in polar liquids.

ELECTRUM. Electrum is a native alloy of gold and silver in which the latter metal may be present in quantities up to 40%. Electrum from the Urals is said to carry 20% copper. The color of electrum is a pale yellow or yellowish-white and the name is derived from the Greek word mentioned in the "Odyssey," meaning a metallic substance consisting of gold alloyed with silver. This same word was also used for the substance amber, doubtless because of the pale yellow color of certain varieties.

ELEMENT (Graph). A *circuit element* is an element of a graph G which is contained in some circuit of G. A *noncircuit element* is an element of a graph G which is not contained in a circuit. The removal of a noncircuit element from a connected graph G leaves G unconnected. The removal of a circuit element leaves the connectivity and the number of vertices invariant. An *oriented element* of a graph is an element with an orientation assigned by ordering the vertices of the element. If the vertices β_1 and β_2 of element ϵ_1 are ordered as (β_1, β_2), ϵ_1 is said to be oriented away from β_1 and toward β_2. See also **Graph (Mathematics).**

ELEMENTARY PARTICLES. See **Particles (Subatomic).**

ELEPHANT (*Mammalia, Proboscidea*). Like most orders of hoofed animals, excepting the even-toed ungulates, Elephants or Proboscideans (order *Proboscidea*) are dying out and are past their zenith. During the Tertiary and the Ice Age there were many proboscidean species, distributed almost throughout the world. They are now represented by just two genera, each with one species, the last survivors of a large group related to hyraxes and sirenids that developed from primitive ungulates in the Lower Tertiary and have evolved as an independent line since the Eocene (50 million years ago). Although elephants are highly developed animals, they seem, with justification, to be primitive organisms. Their extinct predecessors were among the most distinctive mammals of earlier periods of the earth's history.

Trunk. The most striking feature of elephants, besides their great size, is their trunk, which not only substitutes for a hand as a grasping instrument but is also used for touching objects and for smelling. The trunk develops from the upper lip and the nose, and differs structurally in the two extant elephant species. See Fig. 1. Having arisen from the nose, the trunk of course functions in perceiving odors and in respiration, the latter being particularly evident when elephants swim. However, the trunk's most important function is to obtain food and water. When an elephant drinks, it sucks about a bucketful of water 40 centimeters (16 inches) up its trunk, closes the tip of the trunk with the fingerlike structure at the tip, and squirts the water into its mouth. The trunk is also a formidable weapon. It is likewise a highly sensitive tactile instrument, as we see when an elephant picks up a coin from the floor or tugs great loads about. Large motor-nerve fibers extend up and down the trunk, corresponding to the pyramidal tracts in humans.

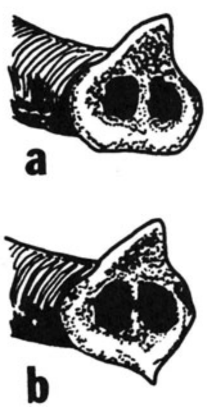


Fig. 1. Tip of the trunk of the (a) Indian elephant, (b) African elephant.

Skull. The powerful skull bears huge ears, especially in the African elephant. The skull bones have spongelike parts, which makes them weigh less than they would if they were solid bone. They are lined with mucous membranes like those found in the nose. The spongelike system reduces the massiveness of the skull in favor of increased mobility, an important compromise since with increasing age the large tusks would become useless if the skull were extremely heavy. The tusks are not elongated canine teeth. They are actually modified upper incisors, and they grow continuously.

Tusks, Teeth, and Ivory. Even at birth, elephants have so-called "milk tusks," which can be up to 5 centimeters (2 inches) long. Permanent tusks erupt when the animal is one year old. One third of each tusk is embedded in the upper jaw bones. Since they are incisors, they have a substantial enamel covering, and over a period of years they develop a strong, sharp tip within which the dental cavity is completely filled. Ivory is a mixture of dentine, cartilaginous material, and calcium-salt deposits on the incisors. The other teeth are also quite distinctive structures. Each of the four jaw halves bears six molars, but not simultaneously. No more than two of these molars are present in a jaw half at any one time. Newborn elephant calves possess the first and second molars; the first is as large as a match box, while the second is about the size of a cigarette pack. The molars wear down as a result of chewing action, and since they also rub against the front of the jaw, they break off in lamellae (in layers). The first molar disappears at the age of 3 to 4 years, while the second molar remains for 6 or 7 years. The third molar becomes functional between the age of 3 to 13, the fourth between 6 and 26 years, and the fifth between the age of 16 and 43. The last molar, which is the size of a small brick, begins to erupt when the animal is 33 years old.

Ivory as already mentioned, is largely the dentine of teeth, a material similar to bone but harder and of different minute structure. It is deposited outside of the layer of cells that produce it in the form of small tubules extending toward the outside of the tooth. The chief sources of ivory for commercial purposes are the tusks of various animals. Elephants' tusks have only a little enamel at the tip and are solid ivory except where the pulp cavity invades the base. The tusks of walruses have also been an important source of ivory, although they are inferior to elephant ivory. The material is used extensively for carved ornaments.

Species Endangerment

Because of the value of their ivory, for many years elephants have been slaughtered by the multi-thousands. Even after the Convention on International Trade in Endangered Species Pact was signed in 1990, by 105 of 110 participating nations, poaching continues at a rampant rate. Malawi and Zambia refused to sign. On the claim of having ample healthy herds, South Africa, Zimbabwe, and Botswana also refrained from signing. The latter three nations have had extensive conservation programs for a number of years. Also, their argument stressed that the ban would only result in rising values for ivory and thus encourage poaching. Essentially, this has been the case in recent years. Estimates indicate that the number of elephants killed for their ivory doubled between 1979 and 1988. Because counting wild elephants is indeed extremely difficult, statistics must be approximate at best. But some experts in the field estimate that the total elephant population, including Africa and Asia, shrank from 1.3 million in 1979 to 608,000 in 1989.

Prior to the 1990 ban, the majority of elephant tusks were shipped to Hong Kong and Japan for processing into jewelry, figurines, and piano keys. Advocacy programs to discourage the purchase of ivory items have enjoyed some success by way of public education. Nevertheless, poaching continues at a high rate. Numerous policing techniques have been proposed and developed, but have enjoyed only marginal effectiveness. For example, DNA fingerprinting has been under investigation by J. C. Patton (Washington Univ.) and others. In recent years, poaching has been extant in Kenya and Tanzania.

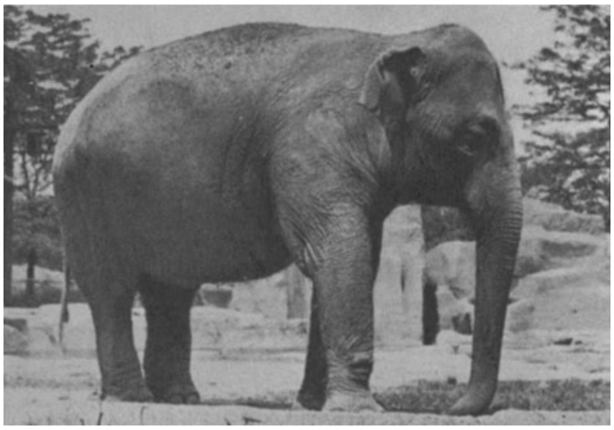
Another major factor in the elephant's plight is human population pressures on the elephant's natural environment.

Elephant Species

The African elephant is the largest of the existing terrestrial animals. The Indian elephant is the third largest, being somewhat smaller than the African Ceratothere or White Rhinoceros. The elephant is characterized by massive structure and by the elongation of the nose and upper lip to form a long prehensile proboscis or trunk. The two upper incisors develop into long tusks in the male and the broad grinding molar teeth grow into position gradually as they are worn off during the life of the animal.

Two species of elephants are recognized: The Indian (*Elephas maximus*); and the African (*Loxodonta africana*). The Indian elephant averages 8 to 9 feet (2.4 to 2.7 meters) in height, but the record is 10 feet, 8 inches (3.2 meters). The African elephant averages about 10 feet (3 meters) in height, with a record of 12 feet, 8 inches (3.8 meters). Adult elephants weight as much as 6 tons (5.4 metric tons). An African elephant may consume up to a half-ton (0.45 metric ton) of grass, twigs, leaves, and fruit in a single day. The African elephant has much larger ears of the two species and the tusks of the African beast are also larger and heavier. See Fig. 2. Numerous extravagant claims have been made from time to time concerning the size of elephant tusks. Reasonable records indicate that the largest tusk from an Indian elephant measured 8 feet, 9 inches (2.6 meters), versus a dimension of 11 feet, 5 inches (3.5 meters) for an African elephant. The tusks weighed 161 and 293 pounds (73 and 133 kilograms), respectively. It is interesting to note that ivory from extinct Mammoths, predecessors of modern elephants, has been dug from the frozen tundra of Siberia and offered to the market.

The life span of the elephant is estimated at about 60 years, although records indicate that some specimens have lived longer. The elephant is fully grown at 25 years, but may breed at 15 to 20 years. The gestation



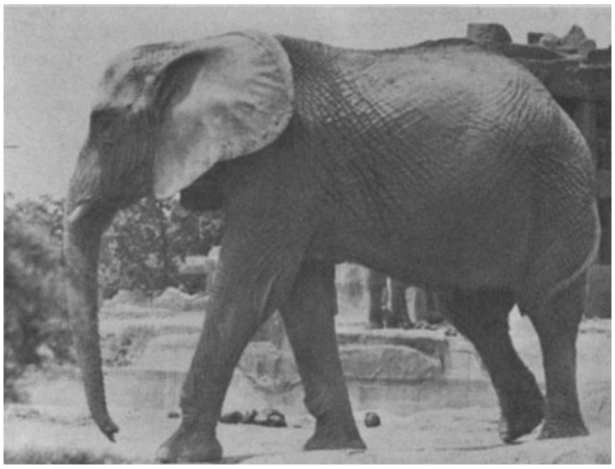


Fig. 2. Elephants: (*Above*) Indian; (*below*) African. (*A. M. Winchester*.)

period of the elephant is from 20 to 21 months. The young are weaned at 5 years. The average speed of an elephant is less than 15 miles (24 kilometers) per hour, although it can achieve a charging speed of 17 to 20 miles (27 to 32 kilometers) per hour. The elephant tends to socialize in herds with 100 or more beasts ambling slowly through their habitat. The animal, because of its tremendous capacity for food, must spend a great deal of its time in browsing and foraging. The tusks are used for fighting and can cause terminal destruction to most would-be attackers. The Big Cats are the elephant's main natural threat, particularly in catching and killing baby elephant calves. Over the years, the elephant has steadfastly resisted domestication.

The mother of an elephant calf is not the only individual to care for the infant; other group members join in, their assistance beginning even before the birth of the calf. Elephant calves are sporadically suckled and fed until the end of their second year, but they are capable of taking solid food shortly after birth. However, they are not able to masticate food at this time. Older herd members help them by collecting and cleaning grass, breaking off branches for them, and cutting the food into small pieces. Young elephants even snatch half-chewed food from the mouths of adults. Newborn elephants sleep more frequently and for longer periods than adults. They often interrupt their play fights to lean against each other and sleep. Since elephant calves are curious and playful, they do not pass up any opportunity to observe nearby buffaloes or deer or to chase a hare, a monitor lizard, or another small animal. Their behavior often creates difficulties for the adults.

Elephant mothers cannot continually watch over their youngsters, but this problem is overcome by keeping the young in "kindergartens," groups of infants that use the best feeding grounds in an area. The group is guarded by different adults. While most of the adults are off feeding, the "baby-sitter" insures that the young stay together and do not run off from the group. Baby-sitters also guard the young while they sleep and pre-chew their food. The baby-sitter role is apparently an unpopular one and is usually assumed, rather involuntarily, at water holes shortly before the group breaks up. Elephant infants have to be forced out of the water, since they tend to splash about and play with each other interminably when they are in the water. The lead cow usually begins moving off as the other mothers and half-grown animals try to get the youngsters out of the water. However, as the lead cow departs, other potential "baby-sitters" begin leaving the young, and the last cow to be with the infants assumes the role of baby-sitter.

Elephant's Sensory System

The extremely keen hearing of elephants has long impressed hunters and mahouts, but it was not scientifically studied until 1951. An elephant could distinguish one specific tone from six pairs of pure tones. In one case the various tones were just one note apart from each other. Elephants also learned simple melodies and rhythms that could be recognized independently of the instrument creating the music (i.e., whether the music was performed on violin, piano, organ, or xylophone).

Elephant cows call their young by slapping their ears against the head. When companions meet, they softly "peep" and "rumble." These greetings sounds are apparently understood by mahouts, too. Elephants trumpet when they are surprised by predators or humans, and the shrill tones of a trumpet often initiate either fight or attack by the elephants. When they threaten, they often beat their trunks against the ground, producing a sound like that of automobile tires striking against a hard surface. Vocalizations either originate in the larynx (as in rumbling and roaring sounds) and are then amplified in the air columns of the trunk, or they are created in the trunk itself (trumpeting).

Olfactory signals undoubtedly play an important role in mutual recognition and in communication. When two elephants greet each other each one touches temporal and cheek glands and the mouth and genital region of the other. We do not know much about how important vision is. For a long time it was thought that elephants have poor eyesight, since their eyes are so small in proportion to the huge body. Furthermore, the field of vision is oriented downward and to the side in the normal head position, causing the elephant to have to lift its head to see in front. However, laboratory experiments have shown that the elephant's vision is as good as that of a horse. It is less able to adapt to changing illumination than humans are.

An observer of a procession of elephants definitely will detect that members of the group are communicating with each other, but at a sound pitch below the human's audible level. Research in Kenya, Namibia, and Zimbabwe indicates that elephants use infrasound for communication. In nature, infrasound is manifested during earthquakes, wind, thunder, volcanoes, and ocean storms, as reported in the Payne reference listed. Prior to some research on elephants, infrasound was not considered as a medium used by animals. This concept was tested by scientists of the Cornell Laboratory of Ornithology by recording elephant sounds at a zoo in Portland, Oregon. Experiments conducted showed that instruments had recorded some 400 calls (infrasound level), which was triple the number of sounds picked up by the researchers' ears. Audible elephant sounds include barks, snorts, trumpets, roars, growls, and rumbles. Part of the latter sounds appear to be infrasonic. Observation of the elephants revealed a fluttering and vibration as air passed through the elephant's nasal passage and the skin on the animal's forehead. This is tentatively considered as the source of the infrasonic vibrations. The researchers suspect that these very low sounds may account for at least some of the behavioral patterns of elephants, such as the female announcing estrus to receptive males.

Diet and Physiology

It seems unusual that the world's largest terrestrial animal makes relatively poor use of its food. Their diet consists of grasses, bamboo, roots, bark, wood, and fruits of specific plants. Some of the most popular items are tender bamboo shoots and the leaves. About half of the food swallowed leaves the body undigested. Elephants spend most of the day preparing and eating their food; adults spend 18 to 20 hours daily with this activity! Their short sleep period of just 2 to 4 hours is sufficient for their needs. Elephants, particularly older animals, often sleep standing up. Sleep is usually interrupted at 15 to 30 minute intervals, during which time the animals check the surroundings for danger. Even in deep sleep they become quickly aroused to any disturbances.

The intestines of an adult elephant exceed the length of those in any other mammal: an elephant has 25 meters (82 feet) of small intestine, 1.5 meters (5 feet) of appendix, 6.5 meters (21.3 feet) of large intestine, and 4 meters (15.1 feet) of rectum. An elephant must typically drink 70 to 90 liters (18.5 to 23.7 gallons) of liquid per day. In the wild and in work the elephant's liquid needs are satisfied during the course of several daily baths. Elephants choose their bathing water carefully, since they also drink where they bathe. Of course, an intake of such huge quantities of water means that elephants urinate impressive quantities. Elephants urinate ten to fourteen times every 24 hours. Wild elephants regularly interrupt their feeding and seek water holes or rivers, both for purposes of drinking and for bathing. They frequently slosh about in mud, which helps cool the body off. When the weather is extremely hot they fan themselves with their ears.

Since elephants are huge animals, they have relatively little surface area for heat loss compared to the internal heat-preserving mass. They only breathe 12 times per minute, and the heart beats 40 times per minute. The average body temperature is 39.9°C (104°F). The blood vessels are quite large: arteries leading to the head have a diameter of nearly 2 centimeters (0.8 inches), and the heart weighs about 12 kilograms (26.5 pounds). The great capacity of the circulatory system may be related to the small sleep requirements of elephants. Elephants in their native habitat tolerate cold better than heat, which they will actively avoid. During the day they avoid open sunny spots, instead withdrawing to the shaded jungles.

Hardly any other animal expends so much time and energy caring for its skin, bathing, massaging, and even powdering it (with dust). The thick skin is not nearly as insensitive as it looks, and it requires constant attention. In young elephants the skin is gray to blackish; with increasing age a pink-white appearance develops, beginning at the base and tip of the trunk, along the edges of the ears, the temples, and on the neck. Finally entire sections of skin have this pale color.

Temperature regulation through the sweat glands is apparently ineffective; elephants are unable to work during the midday hours in hot, tropical climates. Asiatic work elephants are usually rested from 10:30 A.M. to 3:30 P.M., since there is too much danger of overheating during these hours. The secretions from the sebaceous glands, at least in captive animals, are apparently of some help in reducing body temperature.

The Indian elephant is tamed for use as a beast of burden and for handling heavy materials, such as timbers. The elephant does not breed freely in captivity, hence wild herds are the source of supply. The animals are trapped in several ways.

Much has been written concerning the lore of the elephant and its behavior. It is well established that the elephant lives in accordance with certain customs. Apparently, sexual dueling is in order among younger males, but only in accordance with strict rules. The elephant is nomadic and apparently follows one of two plans. In the one case, a herd of animals will make a round trip between two locations during the course of a year; or they make what might be termed a "grand circle" tour that may last some ten years. Although traveling in herds is the norm, solitary elephants are sometimes found; or sometimes a very small group of male elephants. The solitary animals may be explained by the presence of illness or crippling, or just plain aging where an older male may not be able to keep the pace. It is known that elephants suffer from arthritis and rheumatism. An elephant may leave the herd for awhile and then return; or temporarily it may join another herd, but apparently the invitation is extended over only a limited time span. Female elephants of all ages generally are part of a large herd. It is well established that other female elephants assist at birth and in the early care of parent and calf.

Much has been written concerning the intelligence and memory of the elephant. The memory is considered exceptional, but not infallible. Like humans, elephants also forget. There is evidence that at least to some degree an elephant understands human language to a point exceeding that of—say—a dog obeying a command. A well-trained elephant can take verbal instruction from other than its regular trainer and thus obviously depends upon language sounds and to some degree of complexity. In other words, to some extent, the exact terminology need not be repeated each time or by the same person to achieve a desired reaction from the elephant.

In general terms, the elephant is nomadic, highly social, quite intelligent and, essentially as a result of this intelligence, quite temperamental, quite varied in personality from one specimen to the next, and extremely specialized, thus placing the elephant high if not at the top in the order of mammals in terms of advancement from the primitive state.

Over the years, there have been rumors of pigmy elephants. None have been found. The closest would be some of the Loxodonts found in parts of the former Congo territory in Africa, although these hairy races are far from the size of a pigmy animal.

At one time, the elephant, the rhinoceros, and the hippopotamus were classified in one group known as *Pachydermata* (thick-skinned beasts). These animals all have been reclassified, the elephant now in its own group *Proboscidea*.

Additional Reading

Booth, W.: "Africa is Becoming an Elephant Graveyard," *Science*, 732 (February 10, 1989).
Chadwick, D. H.: "Out of Time, Out of Space Elephants," *Nat'l. Geographic*, 2 (May 1991).
Cherfas, J.: "Decision Time on African Ivory Trade," *Science*, 26 (October 6, 1989).
Horgan, J.: "Big-Game Forensics," *Sci. Amer.*, 27 (December 1989).
Lewin, R.: "Global Ban Sought on Ivory Trade," *Science*, 1135 (June 9, 1989).
Payne, K.: "Elephant Talk," *Nat'l. Geographic*, 264 (August 1989).
Ristau, C. A., Editor: "Cognitive Ethology: The Minds of Other Animals," Erlbaum, Hillsdale, New Jersey, 1991.

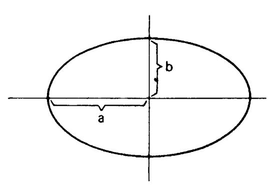
ELEPHANTIASIS. See **Filiarasis.**

ELEPHANT TREE. A relatively large, very noticeable tree found in the desert region of Baja California. Called *copalqum* locally, the tree (*Pachycormus discolor*) has a smooth, light-colored bark reminiscent of the trunk of an elephant. Desert winds constantly prune the tree, this erosion frequently resulting in craggy, gnarled, and grotesque shapes. Foliage consists of small, compound, and somewhat pubescent leaves. Leaf production occurs after intervals of rain. Abundant, small pink flowers borne in panicles give the tree a slightly soft rose coloration when in full bloom. Sometimes, particularly during dry seasons, the tree will flower in the absence of any leaves on the branches. The elephant tree tends to prefer growing in ancient lava flows, where it becomes the predominant plant.

The elephant tree is just one of several thousand plants and trees that are found on the Baja Peninsula. The topography ranges from sea level to elevations of 10,000 feet (3050 meters) or more. Rainfall ranges from about 30 in. (76 cm) in the mountains to very small amounts in the desert areas. The largest elephant tree recorded has a diameter of 30 in. (76 cm) and a height of 30 feet (9.1 meters).

ELK. See **Deer.**

ELLIPSE. A conic section obtained by a cutting plane parallel to no element of a right-circular conical surface. It is the locus of a point which moves so that the sum of its distances from two foci is a constant. Its eccentricity is less than unity. The standard equation may be taken as $x^2/a^2 + y^2/b^2 = 1$. The curve is a central conic for it is symmetric about both the X- and Y-axes. When placed in its standard position, the center of the ellipse is the coordinate origin, the major axis of length $2a$ is along the X-axis, and the minor axis of length $2b$ is along the Y-axis. The distance from the center to either focus is $\sqrt{a^2 - b^2}$; the eccentricity e is given by $ae = \sqrt{a^2 - b^2}$; the length of the latus rectum is $2b^2/a$; the equations for the directrices are $x = \pm a/e$. The distance from any point on the ellipse to a focus is a focal radius and the sum of two focal radii equals $2a$.



Area of ellipse.

If the semi-major axis equals the semi-minor axis ($a = b$), the ellipse degenerates into a circle.

The polar equation of an ellipse is

$$r = \frac{a(1 - e^2)}{1 - e \cos \theta}$$

and its parametric equations are $x = a \cos \phi, y = b \sin \phi$. The equation for the evolute of an ellipse is $X^{2/3} + Y^{2/3} = 1$, where $X = ax/e^2$, $Y = by/e^2$. It is similar in shape to an asteroid and sometimes called that.

With reference to the accompanying diagram, the area of an ellipse $A = \pi ab$. Circumference $C = 4aE(k)$, where $k = 1 - (b^2/a^2)$ and $E(k)$ is the complete elliptic integral of the first kind. This is an approximation for the circumference $C = 2\pi\sqrt{(a^2 + b^2/2)}$.

See also **Asteroid (Mathematics); Conic Section.**

ELLIPSOID. A central quadric surface, given in its standard form with center at the coordinate origin, as

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

where a, b, c are the semi-axes. Sections parallel to each of the coordinate planes are ellipses.

If two of the axes become equal, the surface is a spheroid, which can be generated as a surface of revolution. Consider an ellipse in the XZ-plane $x^2/a^2 + z^2/c^2 = 1$, with $a > c$, so that its major axis is along the X-axis of the coordinate system and its minor axis along the Z-axis. There are then two possibilities: (1) rotate the ellipse about its major axis and the result is a prolate spheroid with $a > b = c$; (2) rotate about the minor axis and obtain an oblate spheroid, $a = b > c$. Sections through the surfaces are circles in both cases: when taken parallel to the plane $x = 0$ for the prolate case: parallel to $z = 0$ for the oblate case.

In the final degenerate case, $a = b = c$, the surface is a sphere.

See also **Spheroid; Surface (Of Revolution).**

ELLIPSOIDAL COORDINATE. A system based on confocal quadric surfaces. If λ, μ, v are the three real roots of a cubic equation in a parameter describing such quadrics, they also locate the position of a point in space, for three mutually perpendicular quadric surfaces intersect at the point. If constants are taken so that $a > b > c$, the surfaces are: (1) ellipsoids, λ = const., $c^2 > \lambda > -\infty$; hyperboloids of one sheet, μ = const., $b^2 > \mu > c^2$; (3) hyperboloids of two sheets, v = const., $a^2 > v > b^2$.

The relation between the rectangular Cartesian coordinates and the ellipsoidal coordinates of a point are

$$x^2 = \frac{(a^2 - \lambda)(a^2 - \mu)(a^2 - v)}{(a^2 - b^2)(a^2 - c^2)}$$

$$y^2 = \frac{(b^2 - \lambda)(b^2 - \mu)(b^2 - v)}{(b^2 - a^2)(b^2 - c^2)}$$

$$z^2 = \frac{(c^2 - \lambda)(c^2 - \mu)(c^2 - v)}{(a^2 - c^2)(b^2 - c^2)}$$

Since x, y, z occur as squares in these relations they give eight points symmetrically located in the Cartesian system. Some convention must then be adopted for the signs of the ellipsoidal coordinates in order to locate a point uniquely.

See also **Coordinate System.**

ELLIPSOMETER. See **Photometers.**

ELLIPTIC CYLINDRICAL COORDINATE. A degenerate case of ellipsoidal coordinates where the surfaces are: (1) elliptic cylindrical with semi-axes $a = c \cosh u$, $b = c \sinh u$, u = const.; (2) hyperbolic cylindrical with $a = c \cos v$, $b = c \sin v$, v = const.; (3) planes parallel to the XY-plane, z = const. A point in this system has rectangular Cartesian coordinates

$$\mathrm{x} = c \cosh u \cos v$$

$$\mathrm{y} = c \sinh u \sin v$$

$$\mathrm{z} = z$$

and $0 \le u \le \infty$; $0 \le v \le 2\pi$; $-\infty < z < \infty$.

See also **Coordinate System.**

ELLIPTIC GEOMETRY. See **Geometry.**

ELLIPTIC INTEGRAL. Any integral of the type

$$\int f(x, \sqrt{R})\, dx$$

where f is a rational function of its two arguments and R is a third or fourth degree polynomial in x, with no repeated roots. It may be reduced, by suitable change of variable, to a sum of elementary integrals and one or more of the following types:

$$u_1 = \int_0^x \frac{dt}{\sqrt{(1 - t^2)(1 - k^2t^2)}}$$

$$= \int_0^\phi \frac{dw}{\sqrt{1 - k^2 \sin^2 w}}$$

$$u_2 = \int_0^x \sqrt{\frac{1 - k^2t^2}{1 - t^2}}\, dt \int_0^\phi \sqrt{1 - k^2 \sin^2 w}\, dw$$

$$u_3 = \int_0^x \frac{dt}{(t^2 - a)\sqrt{(1 - t^2)(1 - k^2t^2)}}$$

$$= \int_0^\phi \frac{dw}{(\sin^2 w - a)\sqrt{(1 - k^2 \sin^2 w}}$$

These are incomplete elliptic integrals of the first, second, third kind, respectively. When expressed in terms of $t = \sin w$, they are Legendre's normal forms. The constant k ($0 < k^2 < 1$) is the modulus and a is an arbitrary constant. If $\phi = \pi/2$, the integrals are called complete.

Series evaluation of the elliptic integrals may be made and numerical tables for them are available. They are called elliptic because they were first studied in order to determine the circumference of the ellipse. Their properties are best studied in terms of their inverse functions. See **Jacobi Elliptic Function; Theta Function; Weierstrass Function.**

ELM TREES. Of the family *Ulmaceae* (elm family), there are close to fifty species of elm trees, notably of Europe and North America. It is estimated that 70% of the hedgerow trees in the English Midlands are elms. Of the species of elms, there are numerous hybrids, cultivars, and clones, and thus there is a resulting complexity of nomenclature.

The American elm (*Ulmus americana*), also called the white elm, has a natural range from Newfoundland to Florida and westward to the Rocky Mountains. Many of the choice specimens were created from grafting. The height of the tree normally ranges from 50 to 100 feet (15 to 30 meters), with a trunk from 20 to 30 feet (6 to 9 meters) in a circumference, and spread of about 70 to 100 feet (21 to 30 meters). About one-third of the way up the tree, the trunk usually divides into a number of very stout branches. The bark is grayish-brown and is furrowed. The leaf is $1\frac{1}{2}$ to 3 inches (3.8 to 7.6 centimeters) in length, alternately spaced, and sharp-pointed. The tree is essentially an ornamental shade trade and widely used for plantings in parks and along streets. Of course, in recent years, the elm disease, described later, has caused the removal of large numbers of these trees.

The cedar elm (*U. crassifolia*) is found mainly in the southern states, ranging westward from the Mississippi basin to Texas and parts of Mexico. The tree usually attains a height between 60 to 80 feet (18 to 24 meters) and a trunk diameter in excess of 2 feet (0.6 meter). The record cedar elm is listed in the accompanying table. The tree is characterized by very small leaves, 1 to 2 inches (2.5 to 5 centimeters) in length, considering the other dimensions of the tree. The September elm (*U. serotina*), also known as the southern or red elm, prefers limestone regions and ranges from southern Kentucky westward into Arkansas and southward to the northern parts of Alabama and Georgia. The tree normally attains a height of about 50 to 60 feet (15 to 18 meters) although some specimens become larger. The leaves are 2 to 3 inches (5 to 7.6 centimeters) long and are of a narrower contour than found on most other elms. The slippery or red elm (*U. rubra*) has a rough bark, deeply furrowed, scaly, and of a dark-brown color. The leaf surfaces also are rough. The tree normally attains a height of about 70 feet (21 meters). The top is often formed in a broad, irregular manner. This species ranges from the lower Saint Lawrence River area westward to the Da-

TABLE 1. RECORD ELM TREES IN THE UNITED STATES[1]

Specimen	Circumference[2] Inches	Centimeters	Height Feet	Meters	Spread Feet	Meters	Location
American elm (1991) (*Ulmus americana*)	312	792	100	30.5	91	27.8	Kansas
Cedar elm (1986) (*Ulmus crassifolia*)	102	259	118	36	66	20.1	Florida
Cedar elm (1989)	127	323	100	30.5	44	13.4	Mississippi
Florida elm (1982) (*Ulmus americana* var. *floridana*)	158	401	94	28.7	54	16.5	Florida
Rock elm (*Ulmus thomasii*)	202	513	117	35.7	122	37.2	Mississippi
September elm (1985) (*Ulmus serotina*)	105	267	150	45.8	64	19.5	Alabama
Siberian elm (1991)[3] (*Ulmus pumila*)	226	574	146	44.5	112	34.2	Michigan
Slippery elm (1989) (*Ulmus rubra*)	240	610	100	30.5	117	36.3	Ohio
Winged elm (1991) (*Ulmus alata*)	185	470	97	29.6	78	23.8	North Carolina

[1]From the "National Register of Big Trees," The American Forestry Association (by permission).
[2]At 4.5 feet (1.4 meters)
[3]Introduced

kotas and Nebraska and south and southwestward from western Florida to Texas. The mucilaginous inner bark has been valued as a home medicine of demulcent qualities. The winged elm (*U. alta*), also called Wahoo elm or cork elm, is smaller than most elm trees, with an average height of 40–50 feet (12 to 15 meters). The tree can grow to much larger dimensions. The bark is of a light-gray-brown color and is close and fine, with perpendicular ridges. The leaves range from 1 to 2 inches (5 centimeters) in length, usually very narrow with sharp points, and of a deep olive-green color. The foliage usually is reasonably dense. The tree ranges southward from Virginia to western Florida and westward to southern Indiana and Illinois and into Texas.

Other species of elm include: the rock elm (*U. thomasii*); the Florida elm (*U. americana* var. *floridana*); the Cornish elm (*U. angustifolia cornubiensis*), a large tree of the British Isles and France; the smooth-leaved elm of Europe and north Africa (*U. carpinifolia*); the Wych or Scotch elm of Europe and north and western Asia (*U. glabra*); the Camperdown, tabletop, Dutch, and Belgian elms, all relatives of *U. glabra*; the Chinese elm (*U. parvifolia*) of northern and central China, Korea, Japan, and Taiwan (disease-resistant); the English elm (*U. procera*); the dwarf or Siberian elm (*U. pumila*); the Jersey or Wheatley elm (*U.* × *sarniensis*); and the Huntingdon or Chinchester elm (*U.* × *vegeta*).

In recent years, the disease-resistant Chinese elm has been planted in the western United States. Its branches are long, delicate, and drooping and it makes an excellent park and street tree, especially for certain parts of California.

The characteristics of record elms in the United States are given in Table 1.

Dutch Elm Disease. The elm for many decades was a favorite for urban planting. Millions of these trees lined the streets of small villages, towns, and large cities. The grandeur of the tree is shown by the accompanying figure. It has been estimated that in 1930 over 77 million elms were located within incorporated areas. It is attributed that Dutch Elm Disease was first noted in Cleveland, Ohio in 1930, when a relatively few trees displayed yellowing leaves in their upper branches, after which they mysteriously wilted and died. Because at first the disease spread slowly, the concern was not great. However, by 1976, approximately 54% of the elms in the United States had died, at which time the population was reduced to about 43 million trees. In New England, the loss of 75% was above average. A map showing the progression of the disease from Ohio in all directions is given in an informative article on "Dutch Elm Disease" by G. A. Strobel and G. N. Lanier [*Sci. Amer.*, 55–66 (August 1981)].

It is alleged that the Dutch elm disease was imported into North America, certainly unintentionally. when it was present in a shipment of elm logs. In early America, wood of the American elm and of the rock elm was used extensively for ship building because it does not splinter. Also the wood bends well, but retains its strength, important characteristics for constructing a wooden ship. The fungus (*Ceratocysis ulmi*) carried by the elm-leaf beetle (*Galeruca scanthomelaena*) and the long-

American elm located at Dundee, Kentucky. (*Kentucky Div. of Forestry.*)

TABLE 2. SOME DUTCH ELM DISEASE RESISTANT ELM TREES

Identification	Developed By
Ulmus carpinifolia × *pumilia* "Urban"	National Shade Tree Laboratory, Delaware, Ohio
U. glabra × *U. carpinifolia* × *U. wallichiana*	(Cultivars are known as Lobel, Dodoens, and Plantyn)
U. laevis Pall	—
U. americana L (NPS 3)	National Capital Region of the National Park Service, Alexandria, Virginia
U. americana "Iowa State"	Iowa State University (Harold S. McNabb)
U. americana "Delaware II"	National Shade Tree Laboratory, Delaware, Ohio
U. japonica "Jacan"	Manitoba Agricultural Experiment Station, Morden, Manitoba, Canada
U. japonica × *pumila* "Sapporo Autumn Gold"	University of Wisconsin (Eugene Smalley)
U. japonica × *pumila* "44-25"	(Sister seedling from the cross that produced Sapporo Autumn Gold)
U. japonica × *pumila* (An unnamed upright European hybrid)	University of Wisconsin (Eugene Smalley)
U. hollandica "Groeneveld 494" (not hardy in upper Midwest)	The Netherlands (Hans Heybroek)
U. davidiana	National Shade Tree Laboratory, Delaware, Ohio
U. Wilsoniana × *japonica* (NPS-5)	(In the collections of the National Capital Region of the National Park Service; and the Arnold Arboretum)
U. × hollandica vegeta (Loud) "Huntingdon"	—
U. hollandica Mill (NPS-8)	National Capital Region of the National Park Service
Unnamed (NPS-36)	National Capital Region of the National Park Service

REFERENCE: "Compendium of Elm Diseases," American Phytopathological Society, St. Paul, Minnesota.

horned beetle (*Saperda tridentata*) grows in the new ring of wood, adjacent to the bark. The tree depends upon this for the flow of sap. In attempting to wall off infection, the tree creates a gummy substance, which ironically clogs tree fluids. Once the sap vessels are blocked, a branch dies and this process is repeated until the total tree is lost. Trees of 100 years or more in age can be destroyed in one season. Various insecticides attack the beetles and the sap-stream also has been injected with the insecticide. Unfortunately the process must be continued and, considering the hazards of effective chemicals, the large number of trees to be treated, and the time and money involved, this is not an ideal approach.

The aroma of rotting tree debris attracts the beetle. This aroma is not emitted by healthy elm trees. The beetles tunnel under the bark where they mate and lay eggs—later to fly away and feed in healthy elms. In so doing, the beetles transmit the fungus to healthy trees.

The most effective response to Dutch Elm Disease, after several decades of research, appears to be to develop resistant species. Well over a dozen such species have been developed, as shown in Table 2, but they may not be easily obtainable. As pointed out by H. V. Wester (National Capital Region of the National Park Service), "Resistant elms are here. The need now is education. Nurseries haven't been growing elms, because people haven't wanted to plant trees that were going to die. But the American elm is the world's finest tree for its purpose. People will soon be planting the tree again with confidence that it will outlive them and their children and grandchildren. There are already the beginnings of nursery preparations to meet the impending demand."

Because resistant varieties are not effective in all areas, biological research continues to better understand Dutch Elm Disease and its vectors. Countermeasures range from the simple to the complex—from beetle trapping and the implementation of sanitation measures to the treatment of trees with fungicides and notably with antagonistic bacteria. *Pseudomonas syringae*, for example, has been used with some success.

One-Tree Forest. Located on a plain of wheat and sunflowers is the Kansas State Forest, which boasts of one tree, the record American elm in the United States. See Table 1. The $1\frac{1}{2}$-acre parcel of land is located in Pottawatomie County in the valley of the Vermillion River, west of the old Oregon Trail crossing. (Four miles east of Wamego on Highway 24, then three miles north on the Onaga Road, and thence one-half mile west.) The state forest was dedicated in 1980. As of 1988, the tree is 272 years old.

ELONGATION (Astronomy). See **Conjunction (Astronomy).**

ELONGATION (Poisson's Ratio). See **Poisson's Ratio.**

ELUTION. In general, a process for extracting a solid substance from a mixture of solids by means of a liquid; as in the recovery of a vitamin adsorbed on an adsorbent by means of a solution. Specifically, a process for the recovery of sucrose from molasses. Quicklime in the proportion of 25% of the weight of the molasses is added, the resulting mass is freed from much impurity by percolating (in "elutors") with 35% alcohol and is then decomposed by carbon dioxide which liberates the sucrose. See **Chromatography.**

ELUTRIATION. The separation of solids by the action of water or other liquids: hence, also the washing of a solid by decantation or a related process.

ELUVIUM. General term for unconsolidated, residual sediments.

EMANATION. See **(Radon.)**

EMBRITTLEMENT. A lowering of the ductility of a metal as a result of physical or chemical changes. Metals may be embrittled under many different conditions. Ordinary steel, wrought iron, and body-centered cubic metals generally, as well as zinc alloys and magnesium alloys suffer a reduction in impact toughness at subnormal temperatures. The effect is only temporary, full recovery of toughness occurring upon return to normal temperatures. Austenitic stainless steels, brasses and bronzes, nickel alloys, aluminum alloys, and lead alloys are not subject to severe embrittlement at low temperatures. Nickel additions to ordinary steels have a favorable effect.

Hydrogen embrittlement of iron and steel may be caused by absorption of atomic hydrogen in electroplating processes or in pickling baths. After such exposure the normal toughness can usually be restored by prolonged aging or a short period of heating at a slightly elevated temperature, as in a steam bath.

Season cracking of high zinc brasses is a severe form of embrittlement resulting in cracking or distintegration. Somewhat similar forms of stress-corrosion cracking occur in many other metals and alloys. Embrittlement of boiler plate, discussed below, may be considered a special case.

Steels and ingot iron may be embrittled by any heat treatment that deposits films of either oxides or carbides in the grain boundaries.

Caustic embrittlement is the development of brittleness in metals such as steel or ferrous alloys, upon prolonged exposure to alkaline substances, like caustic soda, in solution. Failures and explosions in boilers and evaporators have been caused by this action. Effective water treatment essentially has eliminated this condition in boilers.

See also **Corrosion Embrittlement.**

EMBRYO. The developing individual between the union of the germ cells and the completion of the organs which characterize its body when it becomes a separate organism. The term is difficult to limit because some development occurs after birth or hatching and in some species a considerable period of growth intervenes between the completion of the essential structures of the individual and its assumption of separate life. In the latter stage, the organism is called a fetus if it is a mammal; but this term is not applied to the similar period of birds and reptiles. For the embryo in botany, see **Seed.**

At the moment the sperm cell of the human male meets the ovum of the female and the union results in a fertilized ovum (*zygote*), a new life has begun. But before this new life is transferred from the inner, protected life provided by the mother, the new organism will have acquired an age which may vary from a premature 26 weeks to a postmature 46 weeks.

There is no exact procedure for computing the exact age of a child at birth, because the precise date of ovulation of the mother is not established. Although most authorities agree that ovulation occurs about 14 days prior to the beginning of menstruation, it is recognized that such an estimate is hardly more than an average, and that mature eggs (*ova*) may be liberated either sooner or later. Therefore, it is not known how long the new life inhabits the womb before birth, except in rare cases in which the exact occasion of a fruitful coitus is surely known.

But regardless of the length of time spent in the womb by the unborn life, it goes through six preparatory stages before becoming a full-term infant.

The first of these stages is fertilization within one of the Fallopian tubes, which extend from each side of the top of the womb (*uterus*). It is believed that fertilization takes place within the first 24 hours following sexual intercourse. Almost immediately after fertilization of the ovum, the second stage begins. This stage is concerned with the process of cell division. The single-celled zygote becomes a multicelled embryo. The term *embryo* covers the several stages of early development from conception to the ninth or tenth week of life. The early embryo is barely visible without the aid of a microscope; it is considerably smaller than the periods which end the sentence of a typical printed page. The initial series of cell divisions occur as the fertilized egg passes down the Fallopian tube.

Although cell division is still going on in the third stage, when the embryo reaches the womb, the cell cluster has not increased appreciably in size. Up to this time, the embryo is free in the uterus. By the end of the tenth day of development, the fertilized ovum begins to burrow its way into the wall of the uterus. See Fig. 1. This process is known as *implantation*, the fourth stage. It takes about two weeks for the embryo to begin to obtain food from the maternal blood vessels; during this time the developing embryo is probably nourished by the uterine substances it absorbs.

During the fifth stage the growing new life attains an age of eight to ten weeks, has definitive vital organs, as well as partial ability to balance itself within its fluid environment. When these organs are formed, the individual is called a *fetus*. In the sixth stage of prenatal development, the fetus is prepared for and experiences birth, at which time it is called an *infant*, fully capable of existing as a separate entity in the outer world.

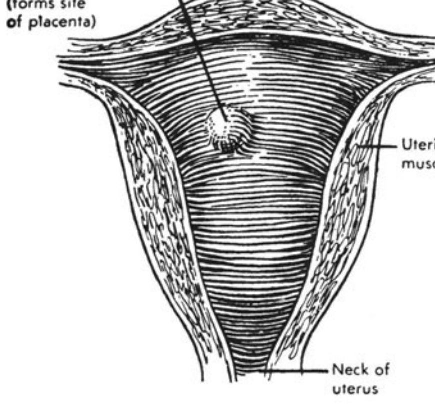


Fig. 1. The human embryo, on about 10th day, becomes embedded in the soft uterine wall. After about two additional weeks, the embryo will derive nourishment through a new placenta which will develop at the site of the attachment.

The period of pregnancy is initiated by the union of the sperm and egg. At the moment of fertilization of the egg (*conception*), a new life begins. The period of pregnancy is also referred to as the period of gestation, and its duration from conception to full-term birth varies between 265 and 285 days in normal situations. See also **Pregnancy.**

Early Days and Weeks of Life. During the fourth stage of development, which coincides with the first two or three weeks after conception, the new life is still not much taller than the capital letters upon this page. It can barely be seen and gives little evidence of its presence in the womb. By the third week after conception the embryo indicates its position in the womb by a small elevation. In the weeks since fertilization of the ovum, the weight of the resulting embryo has increased about 10,000 times, its length about 15 times.

Within the first weeks of growth, the outer layers of embryonic cells are undergoing development and providing nourishment; at this time, too, changes are taking place in the thick disc of inner cells (the *blastodisc*) which gives rise to the embryo proper. The uppermost layer of cells of this disc separates from the remainder to form a cavity known as the *amniotic cavity* and the upper layer thus becomes the *amnion*. The cavity remains filled with fluid throughout the prenatal period so that the developing child leads an aquatic existence during the entire period of prenatal life. The amnion is one of the important membranes covering the developing fetus. In this central cell mass, the outer layer of cells on the underside split off to form what is known as the *yolk sac*. The lowermost cells of the blastodisc becomes the *endoderm*. The embryonic structure now appears to be a flattened disc between two hollow sacs (*vesicles*). The upper of these vesicles is the amnion, and the lower is the yolk sac. The relationship of these structures is shown in Fig. 2.

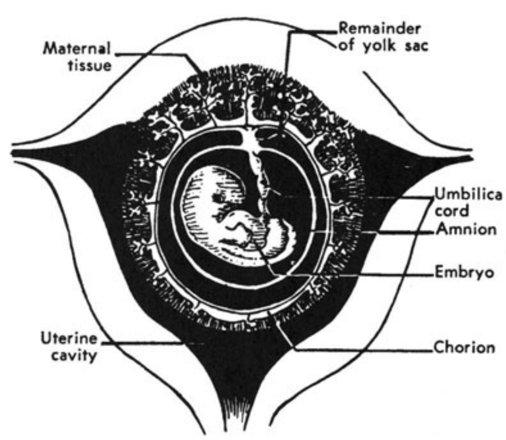


Fig. 2. Protection of developing embryo within the uterus by the chorion and the amnion. The growing child is completely surrounded by amniotic fluid and derives entire nutrition from maternal blood through umbilical cord which is connected to the placenta.

The cells of the embryonic disc segregate to give rise to three main layers of cells which form the definitive organs of the body. The outermost of these cell layers is the *ectoderm*, the middle is the *mesoderm*, and the innermost is the *endoderm*. The middle germ layer, as it is called, rapidly sends out migrant cells which line the entire sac. These mesoderm cells which migrate to the outside of the hollow sphere unite with the external layer to form the *chorion*, a part of which later becomes the *placenta* or the organ by which the fetus obtains food from the mother's blood. Another layer of mesoderm spreads out in a sheetlike fashion to surround the yolk sac. A relatively large amount of mesoderm adheres to the outer membranes to form the body stalk which later will become the *umbilical* cord through which the fetus will receive its nourishment.

Between the second and eighth weeks after conception the three germinal layers differentiate, divide, and combine with each other to lay down the basic body structures, from which the more complicated organism grows. All of the infant body is derived from combinations of the three germinal layers of the embryonic disc. From the mesoderm come supporting structures—muscles, bones, and connective tissues—as well as kidneys, blood and blood vessels, the lymphatic system, and the organs of generation. From the ectoderm are derived the skin and its glands, the hair, the nails, the lens of the eye, the internal and external ear, the mouth and teeth, the mammary glands, the nervous system, and the lower part of the rectum. From the endoderm evolve the respiratory tract, except for the nose; the digestive tract and its glandular outgrowths, including the liver, the pancreas, and the gall bladder; the bladder, and portions of the reproductive organs.

Primitive Streak. During the early part of the third week, the embryonic disc changes shape, so that from the upper surface it appears as an elongated, egg-shaped structure. The cells in the central portion of this disc thicken and form a slight ridge which is known as the *primitive streak*. At one end of this streak, a small knob appears which marks the beginning of the head; in front of this knob the head process forms. The mesoderm is thought by many authorities to give rise to a solid, compact, elongated mass of tissue, the *notochord*. This structure grows forward as well as backward to form the beginning of the backbone. The notochordal tissue continues well into the head region, and on each side of it are formed the bones of the base of the skull.

The primitive streak with its head process divides the embryonic disc into right and left halves. This primitive streak is the first evidence of polarity—cells growing at opposite poles in opposite directions. In the postnatal human animal this polarity of development is clearly evident. An example is the manner in which muscles and tissues grow in opposite directions away from the axis of the spinal cord.

A groove soon courses along the primitive streak, deepens, and presently forms a connecting canal between the amniotic and yolk sac. This is the *neurenteric canal*, forerunner of the *neural* canal. The neural canal is the forerunner of the entire nervous system, including the brain and the spinal cord.

Having organized the area in which will lie the future head of the embryo, the primitive knot shifts, enlarges into an "end bud," and from this bud the lower half of the body arises.

During the third week of prenatal life the embryo is still a tiny organism the size of a large English pea, and the embryonic disc is about the size of the head of a pin. The body now begins to assume a cylindrical form instead of its previously flattened shape. This change is produced by the edges of the embryonic disc growing downward and enclosing the underlying structures. Concomitantly the underlying endoderm rolls itself into a tube which is to form the digestive system, or as it is properly called, the "gut." The mesoderm gathers itself into a number of small segmentally-arranged bundles called *somites*, which later give rise to the deeper layers of the skin and to the muscles and bones. These somites are formed in rapid succession, so that sometimes they are used as an index to the age of the embryo. See Fig. 3.

Differentiation of Primitive Organ Systems. Near the end of the third week, the embryo has formed the beginnings of most of the important organ systems. The anterior or front end of the neural tube closes and the primitive brain starts forming. On either side of the brain, early in the fourth week, is to be found the first sign of the eyes. These are called the *optic vesicles*; they grow out from the brain and appear as bulges on either side of the early head. Back of or posterior to the future eyes are the *auditory vesicles*, which represent the beginning of the ears.

Toward the end of the fourth week of prenatal life, the original three divisions of the brain called the forebrain, midbrain, and hindbrain now become five divisions. By a symmetrical growth the brain makes a series of bends which cause the head region to curve downward with respect to the remainder of the body. The cranial nerves which later innervate the face begin their formation during this fourth week.

The spinal cord, which is formed from the neural tube, becomes thickened on the underside to develop the primitive nerve cells. Previously, when the margins of the neural tube had formed, there were left behind small clusters of cells. These clusters of cells, by uniting with the upper and lower sides of the neural tube, form the spinal nerves. These nerves grow outward from the spinal cord to innervate the organs of the body as they are being formed.

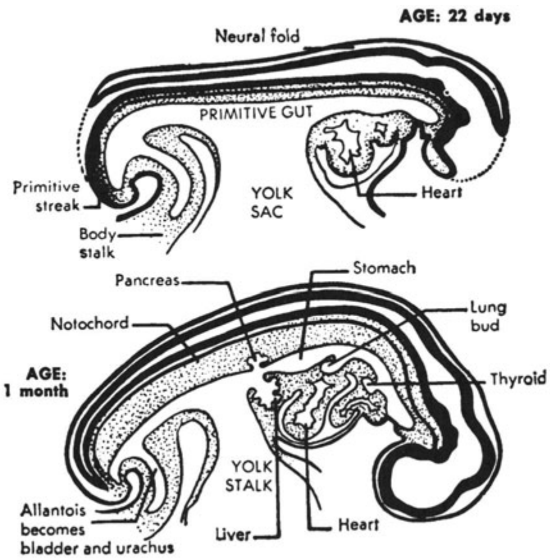


Fig. 3. By end of third week, embryo has commenced to fold over and assumes a more cyclindrical form as contrasted with prior flat form. During this early period, traces of most of the important body structures can first be observed.

The heart is formed by the union of two blood vessels underneath the head. The united tube thus formed grows rapidly in length and bends around itself to form the letter "S." Thus by bending back upon itself the *ventricle*, or main pumping organ of the heart, is formed. Rapidly it changes into an incomplete, four-chambered structure. The heart begins to beat during the third week and continues beating throughout the life of the individual.

The lungs develop from endodermal tissue by forming a bud, which later branches into two lung buds. Each lung bud forms two *bronchi*, from which later are formed the *bronchioles* and finally the *air sacs*.

The primitive gut gives rise to some of the glands such as the thyroid, pancreas, thymus, and parathyroid. From the gut, a pouch grows downward and invades the circulatory system to form the liver. The hind part of the digestive system produces a pouch known as the *allantois* which remains relatively undeveloped in human beings during most of embryonic life. The urinary bladder may be considered a remainder of part of the allantois. The limbs first appear during the fourth week as tiny buds from the midside region and from the hind region of the embryo. They later develop into the arms and legs of the fetus.

Early Fetal Development. After the first eight weeks, the embryo has become a fetus; that is, it now roughly resembles the ultimate adult human being. Prior to this time it would have been impossible to determine by observation whether the embryo was that of a human being, pig, goat, dog, or monkey.

During the third month, there is a rapid growth of the fetus so that by the end of the third month the weight has increased eight- to tenfold. The facial features have shown a marked change; the eyes have migrated inward so that they are no longer on the sides of the head. A bulging high forehead, a small slitlike ear, widely separated nostrils, and a large slitlike mouth characterize the earlier part of the third month's development. The upper limbs show sufficient development so that one may readily discern the fingers, wrist, and the forearm. The lower limbs are relatively smaller and less developed. The liver begins to function during this period. The intestine becomes a coiled structure. At the beginning of the third month the internal organs of reproduction have become sufficiently developed to enable one to distinguish between the sexes. The external genitalia, however, are still in the asexual stage, so that externally both sexes appear the same. Some of the bones are beginning to calcify.

See Fig. 4(a) through (f).

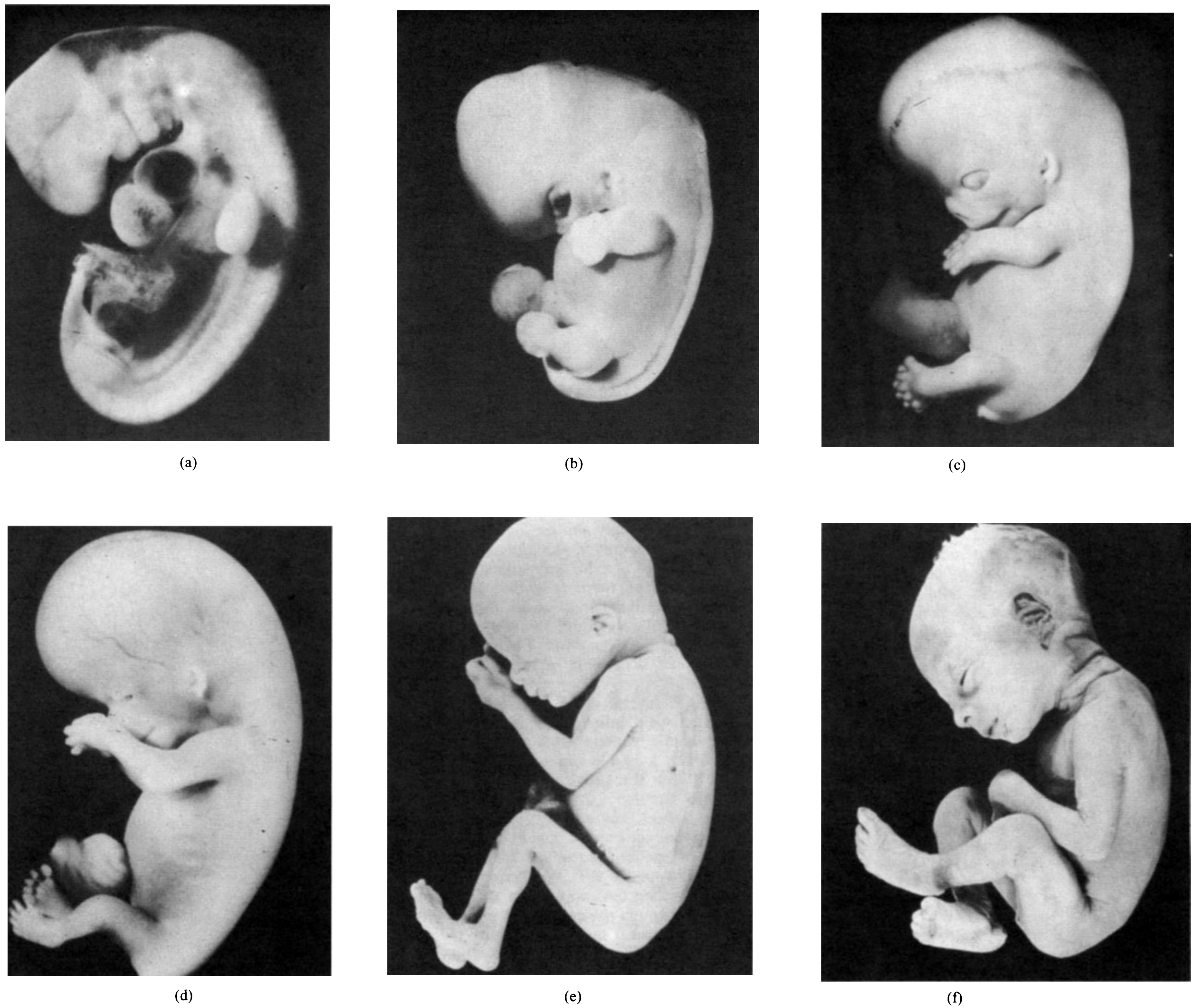


Fig. 4. Various stages in development of human embryo and fetus: (a) At 29 days, magnified 9 times; (b) at 37 days, magnified 4 times; (c) at 42 days, magnified 3.5 times; (d) at 56 days, magnified 2 times; (e) at 4 months (now called a fetus), reduced to about 1/2 to actual size (note disproportionate head size); and (f) at 6 months, shown from 1/2 to 1/3 actual size (note body is now more proportionate to head); (a) through (d) *Carnegie Institution, Washington, D.C.*; (e), (f) *photographed by F. W. Schmidt.*

The Umbilical Cord. The unborn depends for its nourishment, oxygen supply, and the removal of its waste products, upon the mother's blood supply. From the original multicelled embryo, accessory arrangements have developed apace with the growth of the fetus. These, in conjunction with maternal contributions, provide mechanisms to give the embryo nourishment from the mother. From the larger sac which was contained within the covering membrane and surrounded by ectoderm cells, a two-way cord (umbilical cord) develops. This is the connecting structure between the embryo and the placenta. It is attached to the middle of the fetal abdomen.

The Placenta. The original covering membrane (chorion), in cooperation with certain accommodating cells of the womb, evolves the placenta, which is commonly known as "the afterbirth."

Through the placenta, the bloods of mother and fetus circulate independently, and in entirely separate channels. Maternal blood empties into pockets (*sinuses*) in the placenta, from which food materials are absorbed through the thin walls, to pass into the fetal circulatory system. By a reverse process, waste material is picked up by the maternal blood. In addition to providing oxygen and taking up gaseous and fluid wastes from the fetus, the placenta acts as a digestive area for adjusting foodstuffs in the maternal bloodstream to meet the absorption capabilities of the fetus.

The supportive role of the placenta is essential to fetal health and well-being. Besides keeping the fetus alive, the placenta has the additional function of preparing the uterus and birth canal for the delivery. Several hormones are manufactured by the placenta; these are for use sometimes by the mother and sometimes by the fetus. Should the placenta falter in any of its supportive endeavors, the fetus is in trouble.

If the protective functions of the placenta are impaired, noxious products of the fetal metabolism enter and cause disturbances in the maternal blood. The severe and continued vomiting sometimes experienced in pregnancy may be associated with the incomplete functioning of the placenta. In most cases, the placenta prevents infections from reaching the fetus, although sometimes it fails to protect against such diseases as syphilis, smallpox, and German measles. After the birth of the baby, the placenta is expelled from the uterus.

The Protected Fetus. When the placenta and the umbilical cord have been formed, during the first eight weeks of unborn life, the fetus rests within a closed membrane (amniotic sac) which fills the inside of the womb. This fluid-filled sac absorbs shock, equalizes pressure, prevents the fetus from adhering to its protective enclosure, and provides nourishment. After the first 8 weeks, the primitive, but fast-developing muscular and nervous system of the fetus allows it spontaneous movement. Within the amniotic sac, the fetus has room to rearrange its posture. During the seventh week, the middle vestibule of the ear becomes functionally alive. This development allows the embryo to balance itself. The semicircular canals of the middle ear are structures providing for maintenance of static equilibrium throughout life. At birth of the baby, they are of adult size.

Fourth Month to Birth. During the fourth month, there is considerable development of the abdomen so that the head is less out of proportion to the remainder of the body. Hair begins to appear on the head. During this time, the mother becomes aware of movements of the arms and legs.

During the fifth month the lower abdomen and legs become proportionately larger. The legs and arms show vigorous, active movements during this month. A thin silky hair, which disappears during the succeeding weeks, is deposited over the surface of the body. During the sixth month, the fetus increases in size and the organs in complexity. The embryo is lean, with little fat immediately beneath the skin. The skin is protected by a thick, oily secretion of the external glands. Eyelashes and eyebrows are present, and the eyelids have become separate.

The seventh, eight, ninth, and tenth lunar months are characterized by the maturation of the fetus. There is a layer of fat deposited beneath the skin during the last two months of unborn life. This fat protects and nourishes the infant during its early existence in the external world. During these last months before birth, the organs carry on their functions in much the same manner as they will in the external world. The fetus swallows amniotic fluid which passes through the walls of the stomach and the intestine. The kidneys likewise may function slowly and discharge their contents into the amniotic fluid. Rhythmical movements occur in the intestine and the stomach, but their contents are not emptied into the amniotic fluid. During this period the mother's body is active in the elimination of waste material from the fetal body.

In the ninth lunar month, redness which heretofore has been considerable in the fetal skin, now fades. The body becomes rounded; the nails project. Weight is from five and one-half to six pounds. The fetal infant is complete except for the finishing touches which are accomplished in the tenth and last lunar month before birth.

As the fetal body produces glandular secretions and excretions in preparation for changes to be encountered through birth, the body becomes firm, sturdy, and round. By the time the baby is ready to be born, its many body functions—heartbeat, blood pressure, temperature regulation and, as it is being born, its breathing—have been correlated.

Studies of the Fetus

During the last few decades, there has been a great expansion of the knowledge of the fetus. Problems which have confounded obstetricians for centuries are being analyzed and it is becoming possible to treat the new life as a *patient*, along with the mother. This medical discipline is known as *fetology* and has been made possible mainly through the development of new instrumental exploring and measuring techniques.

Amniocentesis. This is the extraction and analysis of some of the amniotic fluid from the sac surrounding the fetus. Study of this fetal fluid yields clues to many obstetrical problems, for example, the complications that arise in the unborn children of mothers who are Rh-negative, diabetic, or hypertensive. An estimate of fetal age can be determined, as well as the sex of the fetus. In amniocentesis, a hollow needle is inserted through the mother's abdomen and the wall of the uterus into the amniotic sac. A small sample of the amniotic fluid is thus obtained. In certain situations, such as Rh incompatibility, a transfusion can be given to the child while in the womb. See also **Blood**. The amniotic fluid contains clues to numerous factors that determine the further development of the fetus and the infant. Because the fluid contains cells from the fetal skin, chromosomal analyses (*karotyping*) can be performed. Potentially dangerous genetic disorders may be diagnosed. Although there is tremendous interest on the part of the obstetrician and, of course, the parents in the state of the fetus, most authorities agree that the procedure should not be undertaken simply out of interest in the sex of the fetus, but rather the technique should be reserved for those instances where family histories or other factors indicate distinct medical advantages to the obstetrician. Although the procedure is generally regarded as safe, it is nevertheless an *invasive procedure*.

Fiberoptic Camera and Endoscope. The *amnioscope* is an illuminated endoscope which can be inserted through the cervix and placed directly against the cervical membrane at any time from the 30th week of pregnancy until delivery. The color of the amniotic fluid will be indicative of whether the fetus is in distress and/or ready for delivery.

The fiberoptic camera is a miniature camera connected to a needle which is inserted into the uterus. Within the needle are fibers that refract light into the lens. Although the instrument yields a picture only one square inch (2.5-centimeters square) in size, it does allow direct observation of the fetus.

Ultrasonography. This is a *noninvasive* procedure and considered by many to be essentially free of risk. The technique can be used to measure the size of the head of the fetus as well as other features. A pulsed beam of ultrasound passing through the fetal head is partially reflected by the skull margins and by the variable density within the brain. A given echo indicates the size of the fetal head. This can be an accurate aid in determination of fetal size and weight. The method can be used after six weeks of pregnancy to visualize the amniotic sac. Progressive ultrasonographic images will indicate the rate of fetal growth.

Maleness. From years of studies directed to understanding human sex-chromosome anomalies, it has been well established that sex determination is mainly a function of the presence or absence of a Y chromosome. The latter is both necessary and sufficient for male development. As observed by Kidd (see reference), exceptions to the foregoing rule have been the main resource for learning what specific genetic information on the Y chromosome is responsible for maleness. Even as early as 60 days of gestation, normal male development appears to require both the inhibition of the "female" (Müllerian) systems and the stimulation of the "male" (Wolffian) systems. These chemical changes are brought about by substances produced by the developing male gonads. Thus, the ultimate cause of sexual differentiation occurs very early, at the time when the gonads begin to differentiate. The search for the gene determining gonadal sex is now focusing on the small euchromatic short arm of the Y chromosome. Cytogenetic observations and DNA studies have implicated this region of the genome. As pointed out by Kiel-Metzger et al. (see reference), the region is sufficiently small and recombinant-DNA methods are sufficiently advanced that the steps of the search are clear. It is believed that within the near future, this entire segment is likely to be cloned. Once cloned, the DNA can be sequenced, and numerous new methods brought to bear—first to look for expressed sequences and then to determine the function of those genes. The ultimate cause of maleness may then be identified, thus providing one more answer to the many questions of human biology being answered by modern molecular genetic techniques.

Vulnerability of Unborn

Extensive studies have been made of the placenta which serves as the interface between the fetal blood supply and the mother's blood. The placenta anatomically separates the circulatory systems of the fetus and the mother. Exchanges of substances take place across this interface. There are, however, important and sometimes tragic exceptions. In the early 1960s, one of the most dramatic examples of how damaging influences from the outside can reach the fetus via the maternal bloodstream was discovered. In December 1962, a tranquilizing and sleep-inducing drug *thalidomide* was taken off the market and all samples were recalled. The drug had not been approved for use in the United States, but had enjoyed distribution in Germany, the United Kingdom, and a few other European continental countries.

This drug, which had been believed to be safe enough even to be given to babies, was causing a rare malformation in infants of mothers who took the drug during the sixth to eighth week of pregnancy, the period during which the limbs are forming. The most common malformation was *phocomelia* or "*seal limbs*," in which the arms and legs were often absent and there were seallike "flippers" in their place. Phocomelia was also often accompanied by internal abnormalities, even some affecting the heart. About one-third to one-half of the babies died within a few days of birth.

Since that time, the medical profession and the drug manufacturers have exercised vigilance in this regard and various regulating agencies in different countries have exercised more aggressive controls over drug testing and approval for distribution, with safe administration during pregnancy being a major concern. Only a few examples of offending drugs can be given here. For example, adenine arabinoside, sometimes used to treat acute herpes simplex keratoconjunctivitis and recurrent epithelial keratitis caused by herpes simplex types 1 and 2, should not be given systemically to pregnant women because of its known teratogenicity and toxicity to the embryo (in experimental animals). Anticoagulants, if used during pregnancy, must be administered with extreme discretion. Although it has been shown that heparin does not cross the placenta and thus does not anticoagulate the fetus, it does increase the potential of maternal bleeding. Warfarin anticoagulants, on the other hand, do anticoagulate both the mother and the fetus. Serious problems have been reported as resulting from anticoagulant therapy during pregnancy, including nasal hypoplasia (incomplete development) and stippled epiphyses (parts of bone; see **Bone**), particularly when warfarin was administered during the first trimester of pregnancy. See also **Anticoagulants.** Anticonvulsant drugs have been shown to increase the rate of malformations in offsprings of women who have been on chronic anticonvulsant drug therapy. There are many other examples. Much greater attention is also being given to the effects of chemicals found in the environment.

The effects of ingesting alcohol during pregnancy are mentioned frequently in the literature. In addition to those drugs which may be prescribed to nonpregnant women with full justification, but which may be contraindicated during pregnancy, there are the so-called hard drugs (street drugs) which poorly informed persons sometimes take. Drug addiction, of course, is an anathema to the pregnant woman. See **Addiction (Drug).** Smoking, a lesser form of addiction, is considered by most authorities in the field as probably harmful in pregnancy over and beyond the effects of heavy smoking on all persons as a precipitating factor of several diseases. The well informed pregnant woman will avoid smoking throughout her pregnancy to assist her body in handling the exceptional demands of that period.

Infections. For many years, rubella (German measles) has been known to be capable of causing birth defects, especially deafness and mental retardation, in children whose mothers had the infection during the first trimester of pregnancy. This usually mild disease can produce devastating effects in the fetus. See **Rubella.**

In addition to the rubella virus, it is now known that other virus infections in the mother may lead to infection in the baby. Depending on the type and severity of the infection, abortion or stillbirth may result, normal development of some of the organs may be prevented (for example, the deafness which often occurs in children of rubella-infected mothers), or the baby may be so infected that its first days or weeks of independent life are an uphill struggle against disease.

Apparently, viruses may infect the fetus at any time from the first few days after conception until immediately before delivery. The incidence of virus infections in fetuses is not known, but it is expected that such infections may account for some otherwise unexplained disorders of the fetus and newborn child.

See also **Embryology; Gonads; In-Vitro Fertilization;** and **Pregnancy.**

Additional Reading

Berkowitz, G. S., et al.: "Delayed Childbearing and the Outcome of Pregnancy," *N. Eng. J. Med*, 659 (March 8, 1990).

Berman, M. C., Ed.: "Diagnostic Medical Sonography," J. B. Lippincott, Philadelphia, Pennsylvania, 1991.

Cherfas, J.: "Embryology Gets Down — to the Molecular Level," *Science*, 33 (October 5, 1990).

Creasy, R. K.: "Preventing Preterm Birth," *N. Eng. J. Med.*, 727 (September 5, 1991).

Ferris, T. F.: "Pregnancy, Preclampsia, and the Endothelial Cell," *N. Eng. J. Med.*, 1439 (November 14, 1991).

Filkins, K., and J. F. Russo: "Human Prenatal Diagnosis," Marcel Dekker, New York, 1990.

Fliescher, A. C., et al., Eds: "The Principles and Practice of Ultrasonography in Obstetrics and Gynecology," 4th Ed., Appleton and Lange, Norwalk, Connecticut, 1991.

Harrison, M. R., Golbus, M. S., and R. A. Filly: "The Unborn Patient: Prenatal Diagnosis and Treatment," W. B. Saunders, Philadelphia, Pennsylvania, 1990.

Hobbins, J. C.: "Diagnosis and Management of Neural-Tube Defects Today," *N. Eng. J. Med.*, 690 (March 7, 1991).

Hollingsworth, D. R.: "Pregnancy, Diabetes and Birth: A Management Guide," 7th Ed., Williams and Wilkins, Baltimore, Maryland, 1992.

Holtzman, N. A.: "What Drives Neonatal Screening Programs?" 802 (September 12, 1991).

Klebanoff, M. A., et al.: "Outcomes of Pregnancy in a National Sample of Resident Physicians," *N. Eng. J. Med.*, 1040 (October 11, 1990).

Mandel, S. J., et al.: "Increased Need for Thyroxine during Pregnancy in Women with Primary Hypothyroidism," *N. Eng. J. Med.*, 91 (July 12, 1990).

Mollier, J. H., and W. A. Neal, Eds.: "Fetal, Neonatal, and Infant Cardiac Disease," Appleton and Lange, Norwalk, Connecticut, 1990.

Palca, J.: "Fetal Brain Signals Time for Birth," *Science*, 1360 (September 20, 1991).

Pearlman, M. D., Titinalli, J. E., and R. P Lorenz: "Blunt Trauma during Pregnancy," *N. Eng. J. Med.*, 1609 (December 6, 1990).

Reece, E. A., et al., Eds.: "Medicine of the Fetus and Mother," J. B. Lippincott, Philadelphia, Pennsylvania, 1992.

Resnik, R.: "The 'Elderly Primigravida' in 1990," *N. Eng. J. Med.*, 693 (March 8, 1990).

Sapienza, C.: "Parental Imprinting of Genes," *Sci. Amer.*, 52 October 1990).

Scarpelli. E. M., Ed.: "Pulmonary Physiology: Fetus, Newborn Child, and Adolescent," Lea and Febiger, Philadelphia, Pennsylvania, 1990.

Tew, M.: "Safer Childbirth?: A Critical History of Maternity Care," Chapman and Hall, New York, 1990.

Volpe, J. J.: "Mechanisms of Disease: Effect of Cocaine Use on the Fetus," *N. Eng. J. Med.*, 399 (August 6, 1992).

EMBRYOLOGY. The science which deals with the development of the individual from the union of the germ cells to the completion of its bodily structure. Although the term embryo cannot be precisely limited, the science of embryology is concerned with all development prior to birth or hatching.

Development of the fertilized ovum begins with the process of cleavage. Following cleavage a process of gastrulation gives rise to two or three germ layers and from this point the development of specialized tissues and organs goes on by gradual steps, all based on the subdivision and differentiation of many cells.

The processes of change by which germ layers give rise to other structures are varied. In some cases masses of cells grow out in solid protuberances from an existing source. This process is called budding and is exemplified by the appearance of legs and other appendages on the surface of the body. Other structures are developed by the pushing in or out of layers of cells. If the new part pushes in the layer it is said to invaginate, and if it pushes out, to evaginate. Hollow organs may also be formed by the splitting of solid masses and parts may separate by splitting from such masses; either process is delamination. A good example of evagination is the pushing out of a blind sac from the embryonic pharynx of vertebrates to form the respiratory system, and invagination is illustrated by the pushing in of ectoderm to form the stomodaeum which becomes the oral cavity in part. The formation of the vertebrate excretory tubules as solid knots of tissue whose cavities arise by internal splitting is a case of delamination.

The details of development of any species or groups of animals are complex. Vertebrate embryology has been worked out in great detail and is fairly uniform, but the number of invertebrate forms is so great that their embryonic development cannot be concisely summarized. See also **Embryo.**

In the vertebrates, once the germ layers are formed their further development is the formation of organs and tissues and in some species extraembryonic membranes, with the exception of the mesoderm. This layer gives rise to diffuse mesenchyme and its compact portions differentiate into three regions, the dorsal, intermediate, and lateral or ventral mesoderm. The first subdivides into two longitudinal series of metameric masses, the mesodermal somites, flanking the middle line of the body where the notochord lies. This skeletal primordium is independently derived from the same source as the mesoderm. The lateral mesoderm splits to form an outer somatic layer associated with the body wall and an inner splanchnic layer which envelops the viscera. The split forms the coelom or body cavity. From this point the mesoderm, like the other germ layers, gives rise directly to organs of the body. The organs and systems developed from each embryonic tissue are listed under germ layers.

In the field of experimental embryology an effort has been made to learn of the controlling factors in development by subjecting embryos and ova to various unusual conditions. By exposure to chemical stimuli, unusual temperatures, radiation, and the effects of centrifuging, many abnormal results have been recorded. It is evident from these results that development, like the life of the organism, is conditioned by a delicate balance of environmental factors. The response of inherited potentialities to this balance in the development of normal organic structure links embryology very closely with the subject of heredity.

Although the word embryology is most commonly used to refer to the development of the embryo of the vertebrate animals, it also refers to the study of the development of embryos of invertebrate animals and plants.

EMBRYONIC FISSION. The subdivision of a single ovum at some stage in its development into parts which give rise to complete embryos. Polyembryony.

As a result of this process a single egg of many insects (parasitic *Hymenoptera*) and of some rotifers develops into several or many individuals.

EMBRYO (Yolk Sac). See **Yolk Sac.**

EMERALD. This beautiful green variety of the mineral beryl has been known since ancient times and always prized as a gem, both because of its color and relative rarity. It is frequently cloudy or flawed, hence the expression "rare as an emerald without a flaw." The original source of emeralds seems to be the so-called Cleopatra's mines in Egypt, where in a range of low mountains about 15 miles (24 kilometers) from the Red Sea, they are found in schists. The quality of these emeralds is not high, but there is much evidence of considerable workings in a former period. See also **Beryl.**

Although emeralds are found in the Urals and to some extent elsewhere the most important locality for emerald is at Muso, Colombia, South America, about 75 miles (121 kilometers) northwest of Bogotá. These mines are believed to be in part at least the source of the emeralds which Cortez and the Spanish conquistadores ruthlessly seized and which were believed for a long time to have come from Peru.

The word emerald is probably derived from the Persian.

EMERY. See **Corundum.**

EMESIS. Commonly termed vomiting, this is the expulsion of the stomach contents through the mouth. This is accomplished by reversal of the direction of the normal waves of peristalsis in the gastrointestinal tract. Vomiting may accompany almost any disease, but in particular often accompanies viral gastroenteritis (intestinal flu), intestinal obstructions, kidney infections, intracranial pressures which may result from blood clots, concussions, meningitis, and tumors, and the ingestion of many offensive drugs and poisons. A person also may vomit when exposed to a particularly offensive odor or terribly unpleasant subject, as upon viewing an accident or other tragic event. Usually vomiting is preceded by nausea, an imminent desire to vomit, accompanied frequently be feeling of weakness, faintness, sweating, vertigo, headache, increased pulse rate, and salivation. Numerous drugs are available to encourage as well as to discourage vomiting.

EMETIC. A substance or drug that induces vomiting, either by direct action of the stomach or indirectly by action on the vomiting center in the brain.

EMF. See **Electromotive Force.**

EMISSION COEFFICIENTS (Einstein). In Einstein's derivation of the Planck radiation formula for a black body three coefficients were introduced which are of importance in the consideration of spectral intensities. Assume two quantum states m and n of a system such that the energy level E_m is higher than E_n; then transition from the upper to the lower state is accompanied by emission of radiation of frequency

$$\nu_{mn} = \frac{E_m - E_n}{h}$$

Assume further that there are a large number of identical systems (atoms or molecules) in equilibrium with black body radiation at a temperature T. Then the rate at which systems pass spontaneously from state m to n, by emission of radiation, is given by

$$-\left(\frac{dN_m}{dt}\right)_1 = A_m^n N_m$$

where A_m^n is known as *Einstein's coefficient of spontaneous emission*, and N_m is the number of systems in state m. But the rate at which systems can pass from the upper to lower state is also dependent upon the density of the radiation, $\rho(\nu_{mn})$, so that there is a second process defined by the relation,

$$-\left(\frac{dN_m}{dt}\right)_2 = B_m^n N_m \rho(\nu_{mn})$$

where B_m^n is known as *Einstein's coefficient of induced emission.*

Furthermore, the system can pass from the lower to the upper state by absorption of radiation, and the rate of this reaction will be given by

$$-\left(\frac{dN_n}{dt}\right)_2 = B_n^m N_m \rho(\nu_{mn})$$

where N_n = number of systems in quantum state n, and B_n^m is known as *Einstein's coefficient of absorption.* See also **Photoelectric Effect.**

EMISSION (Photoelectric). Electron emission from solids or liquids resulting directly from bombardment of their surface by photons. See also **Black Body;** and **Planck Radiation Formula.**

EMMISIONS. See **Acid Rain; Coal; Coal Conversion Processes; Petroleum; Pollution (Air); Wastes and Pollution; Water Pollution.**

EMISSIVE POWER. The emissive power of a body is equal to its emissivity multiplied by the emissive power of a black body at the same temperature. The emissive power of a black body (perfect or complete radiator) is the total radiation from the black body per unit area of radiating surface. See also **Wien Laws.**

EMISSIVITY. The ratio of the radiation emitted by a surface to the radiation emitted by a black body at the same temperature and under similar conditions. The emissivity may be expressed for the total radiation of all wavelengths (total emissivity), for visible light (luminous emissivity) as a function of wavelength (spectral emissivity) or for some very narrow band of wavelengths (monochromatic emissivity). Excepting for luminescent materials, the emissivity does not exceed unity.

EMITTANCE. The radiant emittance of a source is the power radiated per unit area of the surface. This may be either the radiant emittance per unit range in wavelength, the spectral radiant emittance, or its integral over all wavelengths, the total radiant emittance. If the radiant emittance is evaluated by the standard luminosity function, it is called luminous emittance. For a perfectly diffusing surface, the luminous emittance is equal to π times the intensity luminance.

EMPHYSEMA. Distention of tissues by gas or air within the interstices. Emphysema is one of the most common disabling disorders of the respiratory tract, a condition characterized by overdistention of the lungs with air that cannot be expelled. At the microscopic level, the walls of the tiny alveoli stretch and eventually rupture, reducing the capacity of the lungs to exchange carbon dioxide and water. Chronic bronchitis of many years' duration almost always precedes the development of alveolar overdistension. Emphysema is a progressive disease that is most common in males over 40. Its cause is unknown, but excessive smoking and atmospheric pollution may be factors.

Obstructional emphysema is classified professionally as one of the chronic obstructive lung diseases (COLD). Others in the group include chronic bronchitis, bronchiolitis, cystic fibrosis, small airway disease, and Kartagener's syndrome. In obstructive emphysema, the lungs are sometimes called "floppy lungs," i.e., the lungs lack the important property of elastic recoil. The actual physiological cause of the disease is still poorly understood. Some researchers suggest that destruction of lung tissue may result from the actions of proteolytic enzymes. They have found that the plasma of emphysema patients lack certain substances (such as alpha globulin that is capable of neutralizing various proteolytic enzymes, including elastase) which are present in the plasma of a person without the disease. Some researchers have suggested that cigarette smoke may cause increased destruction of elastin by increasing the release of proteolytic enzymes from various lung cells.

X-rays of the chest of patients with emphysema indicate a number of characteristic features of the disease—hyperinflated chest, a low, flat diaphragm, wide interspaces, and small heart.

Symptoms of emphysema include a long history of cough and the raising of sputum, and shortness of breath. The shortness of breath is noticeable first on exertion and later with walking and other daily activities. Bouts of wheezing are not unusual. Weakness, lethargy, anorexia, and weight loss also may be present. In patients with significant emphysema, the chest is hyperinflated and at times fixed in the inspiratory position. On inspiration, the entire rib cage is lifted and accessory muscles of respiration are used. The diaphragm is flattened and hardly moves.

Although changes in the lungs are irreversible, it is possible to give the emphysema patient considerable relief and to increase the functioning capacity of the lungs. The patient should be encouraged to live a moderately active life, but to avoid any exertion which might increase shortness of breath. To control bronchospasm, patients should make regular use of bronchodilator aerosols. Thick and tenacious bronchial secretions can be thinned with sputum liquefiers, and deliberate coughing will help to bring them up. Exercises should be done to strengthen the abdominal muscles and permit more complete exhalation. Manual compression of the abdomen during expiration will aid in elevating the diaphragm. Elevating the foot of the bed will produce similar results.

Oxygen inhalation is sometimes necessary for relief of shortness of breath, but it must be used with caution. The safest method is with the intermittent positive pressure apparatus which produces adequate ventilation and also removes carbon dioxide. A return to normal ventilatory function cannot be expected in patients with symptomatic emphysema. The normal course of events is relentless progression, and therapy is successful if it maintains the status quo or merely slows the downward trend. A patient with mild to moderate emphysema may live a long and comfortable life, provided all the factors producing bronchospasm and bronchial irritation are controlled. In contrast, the patient with severe emphysema has a greatly reduced life expectancy.

EMPRESS TREE. See **Catalpa Tree.**

EMPYEMA. Accumulation of pus in the pleural space, a condition known since the time of Hippocrates. Bacteria may enter the pleural space as the result of bronchopulmonary infections, such as pneumonia, lung abscesses, and bronchiectasis. Empyema may be a complication of thoracic (region between neck and abdomen) surgery. A direct injury to the chest region may be a cause. Rarely a rupture of the esophagus is a causative factor. Traditionally, pneumococcus was the principal infective agent (associated with pneumonia) seen in empyema, occurring in 5–10% of pneumonia cases. Early and effective antibiotic therapy in handling pneumonia cases has greatly diminished this cause. Today, *Staphylococcus aureus* is the major cause. *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, and *Escherichia coli* are also seen with some frequency. However, the possible causative agents are many, including anaerobes, *Bacteroides* species, *Fusobacterium*, and *Mycobacterium tuberculosis*, among others.

Present with empyema are fever, dyspnea, chest pain, and cough. With delayed treatment, weight loss usually will occur. Therapy is usually a combination of antibiotic administration and drainage. In acute empyema, thoracentesis or tube thoracostomy may be employed. Surgical drainage may be indicated in chronic empyema. The chronic form of empyema is less frequently seen because of advances in treatment of patients in the acute stage.

The appearance of the fluid from the pleural cavity in empyema varies. It may be watery (*serous*), puslike (*purulent*), or a combination of the two (*seropurulent*). In deep wounds, in addition to the aforementioned procedures, irrigation (*lavage*) with saline solution may be indicated.

EMU. (*Aves*, *Casuariiformes*, *Dromaiidae*). The Emus are flightless inhabitants of Australian bush steppes. Three subspecies which lived on coastal islands were exterminated in the last 150 years. Ancestors of today's emus lived in the Upper Pleistocene (50,000–10,000 years ago) in Australia. See accompanying illustration.

Externally they resemble the rhea. See also **Rhea.** They are about the same size but much more compact and heavier, weighing up to 55 kilograms (121 pounds). The main shaft and secondary shaft (aftershaft) are equal in length so that every feather appears to be double. The wings are small, and are hidden by the rump plumage. There are three toes. The gut and caeca are shorter than in rheas. Their food consists of fruits and seeds. There is no preen gland.

The emu is a fast runner which can reach speeds of up to 50 kilometers (30 miles) per hour. Surprisingly, it also swims well and with endurance.

Areas inhabited by the emu: (1) Emu (*Dromaius novaehollandiae*), Australia, extinguished in Tasmania and many parts of Australia: (2) black emu (*D. minor*), Kangaroo and King islands, extinct.

Incubation and raising young is the male's task, as in the rhea and cassowary. When two emus are paired they stand next to one another with lowered heads and bent necks. They sway their heads from side to side above the ground. Then the female sits down, the male sits down behind her, and he shuffles up and onto her and finally grasps the skin of her nape with his beak. At the same time he utters squeaking or purring sounds, and finally he runs away while the female remains sitting. The nest is a shallow depression located next to a bush. It is simply made with leaves, grass, and bark, and holds 15–25 eggs which come from several females. Incubation takes 25 to 60 days; the great variability is due to pauses during which the male must leave to feed and drink for shorter or longer periods of time. At 2 to 3 years of age, the young are grown and capable of reproduction. See also **Ratites.**

EMULSION. See **Colloid System; Photography and Imagery.**

ENAMEL. See **Paint.**

ENANTIOMORPHISM. See **Amino Acids.**

ENANTIOTROPY. The property possessed by a substance of existing in two crystal forms, one stable below, and the other stable above, a certain temperature called the transition point.

ENARGITE. A grayish-black or iron-black orthorhombic mineral Cu_2AsS_4. It is an important copper ore, occurring in veins of small crystals or granular masses. Often contains antimony up to about 6% and sometimes small amounts of iron and zinc.

ENCEPHALITIS. An infection of brain tissue usually caused by one of several viruses, but which also may be caused less frequently by the ingestion of a drug or toxin, a systemic fungal infection, or by a space-occupying lesion, such as a tumor or subdural hematoma. Postinfectious encephalitis also may occur with mumps, measles, rubella, or chickenpox. The incidence of encephalitis from these infections is no longer common with the exception of rubella, for which no satisfactory vaccine is yet available. The symptoms and course of encephalitis are quite variable, depending upon the causative agent and initial condition of the patient. In encephalitis resulting from some viruses, the symptoms may be a brief illness, so mild that the patient does not seek extra rest; or it may be a grave illness with high fever persisting for several days or weeks. Stupor and weakness of eye muscles are the most notable symptoms in some patients, while less commonly there may be violent delirium, insomnia, and involuntary muscle activity. Muscular rigidity and rhythmical tremor may be seen, as in paralysis. There may be a rapid fatal termination, or the illness may be chronic. Unless the disease occurs during an appropriate viral season for which epidemiologic data are available, diagnosis can be difficult. Specific diagnosis usually requires extensive laboratory examination of blood and cerebrospinal fluid. In the case of enteroviruses, throat and stool cultures may be a useful diagnostic tool. In the case of herpes simplex type 1 encephalitis, a CAT scan will reveal involvement of the temporal, frontal, or parietal lobes. In some cases, open brain biopsy may be indicated for the purpose of cerebral decompression and of ruling out lesions, such as brain abscesses, tumors, or fungal infections.

Simple viral encephalitis usually is self-limiting, running a course of about two weeks. Authorities indicate that less than 50% of the roughly 2000 cases of encephalitis reported in the United States each year are properly diagnosed. The incidence in any given year depends primarily on whether arboviral or enteroviral epidemics occur in any given year. The togaviruses are responsible for most cases of the disease. Other agents, less frequently encountered, include Colorado tick fever virus, varicella-zoster virus, mumps virus, and herpes simplex type 1 virus. Sometimes aseptic meningitis caused by coxsackievirus B presents essentially an encephalitic picture.

See also **Virus.**

ENCODER (Computer System). A device or subsystem which will accept an input and produce an output in coded form. This definition includes digitallogic configurations which convert a digital input word in one code into an output word in a different code. Examples would include a decimal-to-binary or a decimal-to-octal decoding circuit. Encoder also refers to a device or subsystem which converts an analog quantity into a digital representation by the use of a quantization technique. See also **Encoder (Electromechanical).**

ENCODER (Electromechanical). A device that provides position, direction, speed, and displacement information. The rotary encoder, frequently used in automated industrial systems, satisfies the IEEE definition of *encode: "to produce a unique combination of a group of output signals in response to each group of input signals."* Sometimes the comparison of an encoder with the familiar micrometer caliper (Fig. 1) is made. With the caliper, the micrometer screw is turned to accurately measure the dimension of a piece held between the jaws of the device. The micrometer barrel is divided so that measurements to an accuracy of one-thousandth of an inch (0.025 mm) or better can be made. Each revolution of the barrel advances the micrometer spindle 25 thousandths of an inch, thus requiring 40 complete revolutions of the micrometer screw to advance the micrometer spindle one linear inch. Of course, the larger the barrel, the more divisions per turn to read, and consequently the greater the accuracy. Instead of manually reading and interpolating a scale, the encoder translates the simple analog rotation into discrete electrical signals that are directly related to shaft position and hence to the distance traveled. Shaft encoders are of two basic types—*absolute* and *incremental.*

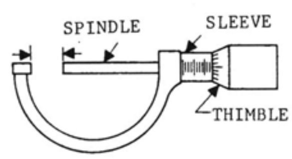


Fig. 1. An encoder, in operation, resembles a micrometer caliper.

Absolute Encoder. This device provides a unique output signal for each single or multiple revolution of shaft gearing. The device outputs a complete binary code (digital output) for each position. Absolute encoders are generally used in applications where position information rather than change in position is important. These devices have an individual digital address for each incremental move and thus the position within a single revolution can be determined without a starting reference. By gearing two or more absolute encoders together, so that the second advances one increment for each complete revolution of the first (reminiscent of a mechanical counter), the range of absolute position can be extended.

The disk type is manufactured with a coded track pattern to provide a digital signal output (0-1 or on-off). The absolute shaft encoder uses either (a) contact (brush), or (b) non-contact schemes of sensing position. The contact type is shown in Fig. 2(a). The device incorporates a brush assembly to make direct electrical contact with the electrically conductive paths of the coded disk for reading address information. The noncontact type, shown in Fig. 2(b), uses optical means (commonly photoelectric) to sense position from the coded disk. In this case, the disk consists of opaque and transparent (to light) segments. These segments are laid down in the same pattern as the electrically conductive paths used by the brush-type encoder.

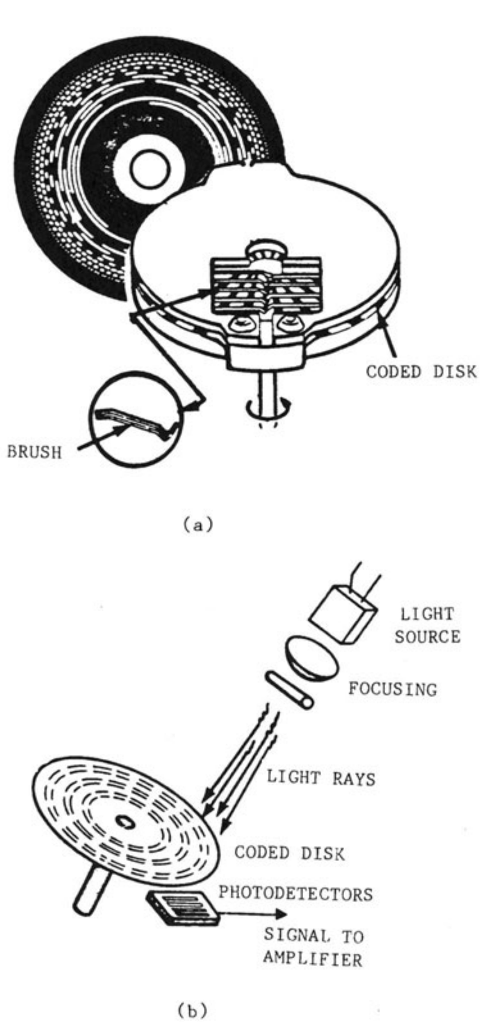


Fig. 2. Absolute encoder: (a) coded-disk type with contact or brush; (b) noncontact, photoelectric type.

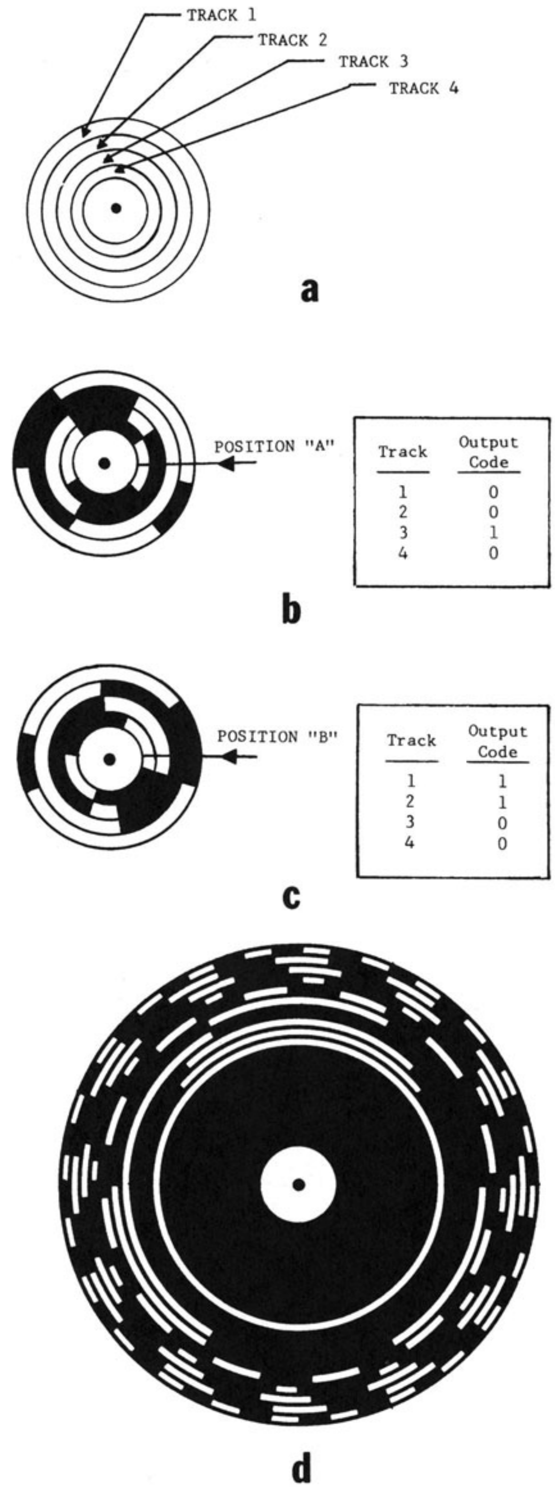


Fig. 3. Operating principle of absolute encoder: (a) Several concentric tracks (only four shown in oversimplified diagram) are present on the disk. (b) Portions of each track are either opaque or transparent to light emanating from a point source. The pattern of opaque/transparent segments is designed so that for every degree of rotation (360°) of the disk, a unique coded address will be presented. The detail required is not shown in diagrams (b) and (c). When the disk is in position "A" as shown in (b), the segments on tracks 1 and 2 are transparent, thus each yielding an output of 0; the segment on track 3 is transparent, thus yielding an output of 1; the segment on track 4 is opaque, yielding an output of 0. Thus, the complete address is 0010. In (c), the disk has rotated so that tracks 1 and 2 are opaque and tracks 3 and 4 are transparent, providing the address of 1100. A reasonable representation of a full disk with 10 tracks is shown in (d).

The principle of operation is demonstrated by Fig. 3. The number of tracks may be increased as well as the segments around the disk until the number of graduations equals the desired resolution. Since position information is directly on the coded disk assembly, the disk has a built-in "memory system" and a system power failure will not cause this information to be lost. Thus, it is not necessary to return to a "home" or "start" position after reenergizing power.

Incremental Encoder. This device produces a symmetrical pulse for each incremental change in position. Pulses from the incremental encoder are counted for each incremental movement from a calibrated starting point in an up/down counter to track position. The operating principle of an incremental encoder (also sometimes called *optical encoder* or *digital tachometer*) is shown and described by Fig. 4. Another area which is application-dependent is the disk assembly. Depending upon the resolution and accuracy, the material used may limit the encoder for some applications. The disk can be made by using slits in metal, or lines on glass. The metal disk is normally a low-resolution device. Glass provides higher resolution and accuracy, but must be handled with care to avoid breakage.

An incremental encoder can be either *unidirectional* or *bi-directional*. A unidirectional encoder yields information about speed or amount of displacement. A bi-directional encoder provides this same information as well as *direction* information, i.e., clockwise or counterclockwise rotation. Direction information is obtained by monitoring two signals electrically separated by 90 degrees. As shown in Fig. 5, phase relationship between these two signals is utilized to determine rotation direction. Incremental encoders may have several tracks. As illustrated, a second track can be used for a "zero" index or "home" reference pulse.

Specifying parameters applicable to incremental encoders include: (1) *Line count*, which is the number of pulses per revolution. The number of lines is determined by the positional accuracy needed for a given application. Standard line counts commercially available range from

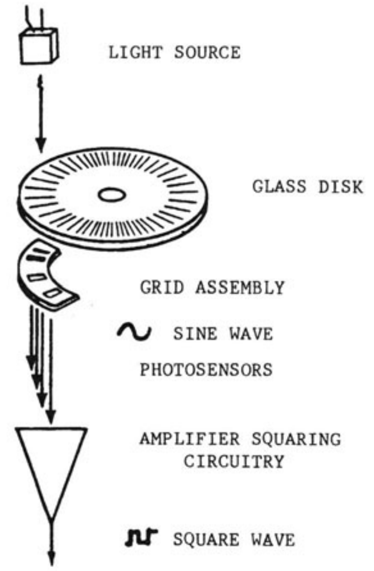


Fig. 4. Incremental encoder. This device provides an output pulse signal as the disk assembly rotates; thus total information is obtained by counting pulses. The disk is manufactured with opaque lines which are aligned with a grid assembly. A light source and photosensors complete the assembly. Light from a light emitting diode (LED) or tungsten filament lamp passes through the transparent segments of the disk and is sensed by photosensors. As the disk assembly rotates, an alternating light/dark pattern is produced. The output from the photosensors is a sinusoidal wave which can be amplified in some situations. Electronic processing transforms this signal into a square-wave pulse for digital circuitry compatibility.

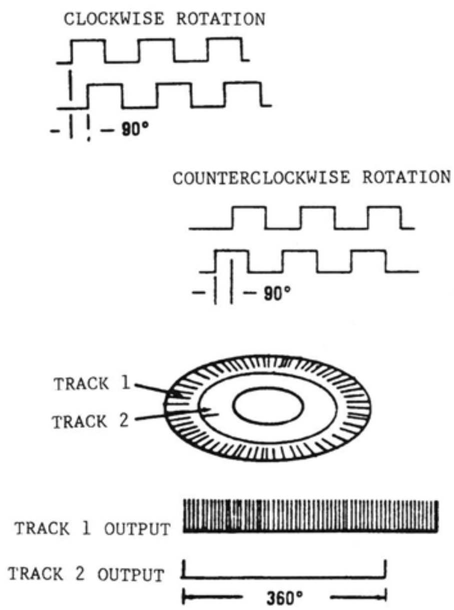


Fig. 5. Top of view shows phase relationship between two signals for determining rotation direction. Incremental encoders may have several tracks. As shown here, a second track can be used for a "zero" index or "home" reference pulse.

100 to 1000 pulses/revolution. In some specially designed "self-contained" encoders, higher line counts are obtainable. (2) *Output signal* can be either sine or square-wave. (3) *Number of channels*. Either one or two channel outputs can be provided. The two-channel version provides a signal relationship to obtain motion direction. In addition, a zero index pulse can be provided.

A typical servo application using digital feedback is shown by the block diagram of Fig. 6. The input command signal loads an up/down counter. The number of pulses in the counter represents the position the load must be moved to. As the motor accelerates (Fig. 7), the pulses emitted from the encoder continue at a faster rate until motor run speed is obtained. During run, the pulses are emitted at a constant frequency directly related to motor speed. The counter counts down to "zero" and, at a determined position, the motor is commanded to slow down. This is to prevent overshooting of the desired position. When the counter is within one or two pulses of the desired position, the motor is commanded to decelerate and stop. The load should now be in position.

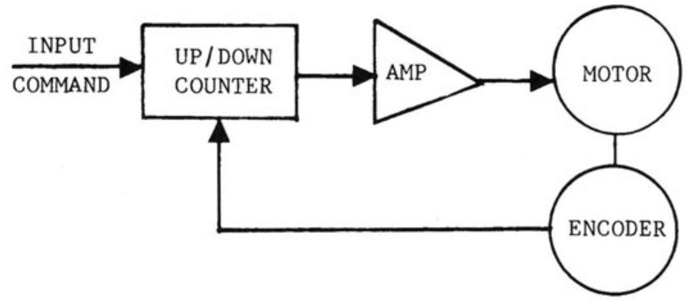


Fig. 6. Block diagram of servosystem using digital feedback to an up/down counter.

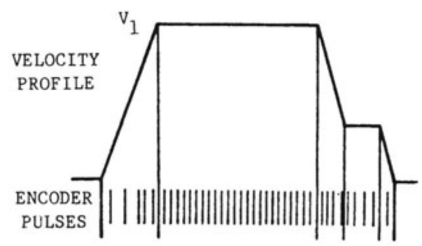


Fig. 7. As the motor accelerates, the pulses emitted from the incremental encoder (digital tachometer) continue at a faster rate until motor run speed is attained. During run, the pulses are emitted at a constant frequency directly related to motor speed. The counter counts down to "zero" and, at a determined position, the motor is commanded to slow down—to prevent overshooting the desired position. When motor is within two pulses of desired position, the motor is commanded to decelerate and stop.

Representative Problem. An incremental encoder is required on a milling machine to provide a digital readout display. The display must read directly in thousands of an inch. The total travel of the milling machine bed is 36 inches. The travel is regulated by a precision lead screw which moves the milling machine bed $\frac{1}{10}$ inch for every revolution (360°) of the lead screw. Since the display must read directly in $\frac{1}{1000}$ inch increments, the encoder must provide 100 pulses per revolution where each pulse represents 0.001 inch.

Solution. An encoder disk is connected to the shaft of the motor and the shaft is rotated. A pulse train is generated by photoelectric means as previously described. These pulses are fed directly into an appropriate electric counter with digital display. Starting from a known reference position, the operator resets the counter to zero. The operator moves the milling machine bed from the zero position until the number 19.031 is shown on the counter. The operator is now exactly 19.031 inches from the zero position.

In some systems, the number 19.031 is entered on the counter's preset function. When the counter counts 19,031 pulses, it stops the travel automatically. At this position, a hole is bored to a specific depth. An encoder on the *z*-axis of the machine controls the drilling to a specified depth. Add to this an encoder for bed travel on the other axis, plus tape control for the preset functions and sequences, and automated numerical control is the result.

Encoder Interfacing. The square-wave output is derived, as previously mentioned, from electronic processing, or shaping circuitry within the encoder package. The output signal level is nominally 5 V and zero (logic "1" is 2.4 V minimum and logic "0" is 0.4 V maximum). Signal distortion may be a result of cable capacitance (length)—the longer the cable, the more distortion. Beyond 30 feet (9 meters) in length, the signal must be reshaped if reliability is not to suffer. Good shielding must be used to keep noise to a minimum.

The sine wave output normally will be used where the designer performs the signal shaping somewhere else in the system, i.e., other than in the encoder package. Signal levels are typically 50 to 100 mV peak to peak into a 2 k ohm load at 40 kHz. A disadvantage is susceptibility to electrical noise because the signal is at such a low level. Signal cables must be isolated from other ac lines and noise generators. Twisted, shielded wires should be used. Signal reshaping usually is not required for distortion, since the receiver is a signal shaper.

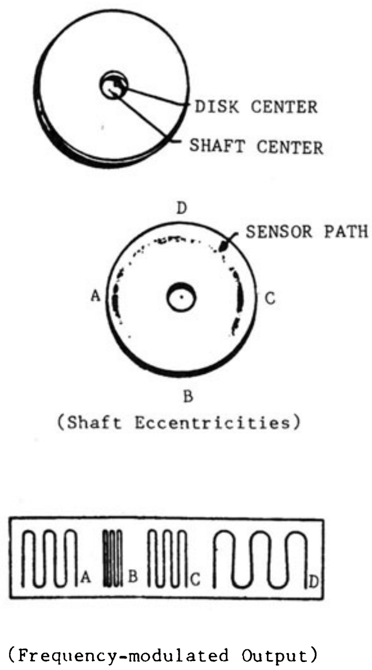


Fig. 8. Shaft eccentricities. If the hub is oversized and/or the shaft is undersized, the encoder will exhibit eccentricities when mounted, thus yielding a moving rather than a fixed center. Eccentric mounting causes a frequency-modulated signal to be superimposed on the encoder output signal. If the sensor reading the line count traces a path on the disk, as indicated here, the resulting output signal will be frequency modulated.

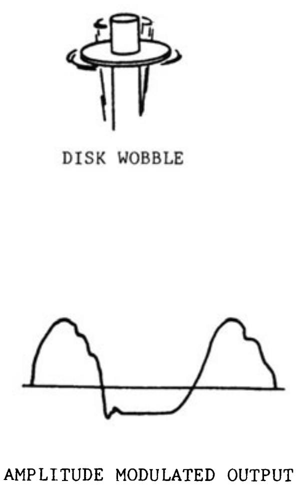


Fig. 9. Disk wobble. If the motor shaft has a large total indicated runout (TIR), the disk assembly will wobble. This will produce an amplitude-modulated signal due to varying illumination received by the sensors. This condition can be corrected either by first mounting the hub onto the motor shaft and machining, or by using dual optical pickoffs.

Distortion is not significant where cable lengths do not exceed 30 feet (9 meters).

In addition to radiated noise, the encoder may be affected by transients caused by its power supply voltage. The regulation is typically specified at ±5% variation, without noise spikes. Spikes, of course, may damage the light source and encoder electronics.

Other Encoder Problem Areas. Of prime importance is the tolerance on the hub inner diameter into which the glass disk is mounted, the motor shaft outer diameter, and the motor shaft's total indicated runout (TIR). If the hub is oversized and/or the shaft undersized, the unit will exhibit eccentricities when mounted and thus yield a moving center rather than a fixed center. Eccentric mounting causes a frequency-modulated signal to be superimposed on top of the encoder output signal. If the sensor reading the line count traces a path on the disk, as shown in Fig. 8, the resulting output signal will be frequency modulated. If the motor shaft has a larger TIR, the disk assembly will "wobble" as shown in Fig. 9. This will produce an amplitude-modulated signal due to the varying illumination of the sensor. This can be corrected by (1) mounting the hub onto the motor shaft and machining, or (2) by utilizing dual optical pickoffs.

See also **Automation; Numerical Control;** and **Position and Displacement Measurement.**

ENDANGERED SPECIES. For a variety of reasons, several important species of life on Earth have been threatened by extinction during the last several decades. Greater awareness of Earth's environment has progressed much over the past half-century, out of which a serious awareness of the effects of environmental change on the existence of various life forms has developed. Polluted air and water affect the natural habitats of most species in one way or other. Other anthropogenically created alterations, such as overfishing, destroying forests and wetlands, also adversely affect some species. Several examples of threatened species are given throughout this encyclopedia. As of 1994, the species regarded as endangered or seriously threatened are listed in the accompanying table. This is only a small part of the total list as developed by naturalists, conservationists, and environmentalists.

ENDANGERED AND THREATENED SPECIES WORLDWIDE

Mammals	
Bear, Baluchistan	Wolf, Gray and Red
Bear, Polar	Zebra, Cape Mountain
Cat, Pardel Lynx	**Reptiles**
Cat, Little Spotted	Crocodile, American
Cheetah	Iguana, Anegada Ground
Chimpanzee, W. African	Python, Indian
Cougar, Florida	Snake, Atlantic Saltmarsh
Deer, Key	**Amphibians**
Deer, Marsh	Frog, Israel Painted
Elephant, Indian	Toad, Mount Nimba
Gazelle, Clark's	**Fish**
Gazelle, Slender-Horned	Catfish, Giant
Gorilla, Mountain	Trout, Cutthroat
Ibex, Walla	**Birds**
Jaguar	Albatross, Short-Tailed
Leopard	Condor, California
Leopard, Snow	Crane, Whooping
Lion, Asiatic	Crow, Hawaiian
Mandrill	Kestrel, Mauritius
Monkey, Long-Haired Spider	Parrot, Paradise
Ocelot	Stork, Oriental White
Prairie Dog, Utah	Woodpecker, Ivory-Billed
Pronghorn, Sonoran	
Rat, Morro Bay Kangaroo	
Rhinoceros, Great Indian	
Sloth, Maned	
Tiger	
Wallaby, Bridled Naitail	
Whale, Humpback	

Source: United Nations Environment Programme, Gland, Switzerland.

ENDEMIC. A term applied to a disease caused by agents (especially of an infective nature) that are constantly present in a particular human community, leading to a generally higher incidence of the disease in that community than elsewhere. See also **Foodborne Diseases.**

ENDOCARDITIS. Inflammation of the endocardium, a thin layer of tissue lining the inner surfaces of the heart. Bacterial endocarditis is a bacterial infection of the endocardium. It accounts for 2% of all organic heart disease. A patient with bacterial endocarditis has an excellent chance for survival with prompt treatment.

Several types of bacteria can cause bacterial endocarditis. It has been known for a long time that bacteria occasionally gain access to the blood vascular system of the body. Usually these invaders are quickly destroyed by the leucocytes, or white cells, of the blood. However, if the bacteria appear in the blood as the result of an infection elsewhere in the body (*septicemia* or blood poisoning), they may be present in very large numbers. Should invading bacteria become attached to the inside of the heart, to one of the valves of the heart, or to the inner wall of one of the major blood vessels, the result is termed bacterial *endocarditis* (affecting the heart); or *endoarteritis* (affecting an artery).

This condition is especially serious because the circulatory tissues are poorly equipped for combating infection. Whereas other tissues of the body may literally wall up an infection so that it can be destroyed by the white cells, the heart and arterial tissues have no such ability to isolate an infection. A large percentage of persons who have bacterial endocarditis have had a previous heart disability. The heart may have some congenital structural defect, or the endocarditis may have resulted from a disease of the heart, such as rheumatic fever. Affected persons usually are young adults, although the disease may attack any age group.

One of the most characteristic signs of bacterial endocarditis is fever, always present in the acute form, but persons with the subacute form may suffer only intermittent fever. The onset of the fever is almost always a result of the presence of free bacteria in the blood stream. The physician may withdraw a sample of blood during a *febrile* period for culture of the organism. The patient also suffers from anemia, which is partly caused by the destruction of red blood cells by the bacteria.

Embolism is also a complication of bacterial endocarditis. Emboli which develop because of the disease may cause *Osler's nodes* in the skin. These are small, raised, reddened areas found most often on the inside of the fingers and toes. They may be somewhat tender, but usually disappear within a few days. Larger and much more painful lumps may appear on the limbs, beneath the skin; usually they remain about a week. Sometimes these are caused by hemorrhage.

When a bacterial embolus lodges within an artery, it may cause a bulging sac from the wall of the artery called a *mycotic aneurysm*. These aneurysms appear in the smaller arteries, such as those that supply the skin. However, they may occur elsewhere.

When emboli become lodged within the blood vessels of the lungs, they produce symptoms similar to those of *hemorrhagic bronchopneumonia*. Emboli affecting the kidneys will cause many of the signs and symptoms of kidney malfunction, but rarely cause fatal *nephritis*. An embolus lodging in the brain may result in widespread damage to nervous tissue by cutting off the blood supply to nerve centers. Probably because of toxins manufactured by the bacteria, the smaller blood vessels (the capillaries) often become unusually fragile. The rupture of the walls of these tiny vessels causes a hemorrhage; the resulting symptoms depend upon the location of the capillaries affected. When capillaries in the skin are affected by the toxins, numerous small, purplish spots appear in the skin. They may be seen almost anywhere in the skin or mucous membrane. When they appear under the nails, the spots often resemble splinters. There may be capillary ruptures on the surface of internal organs, notably the heart and kidneys. In addition to these signs, the spleen usually becomes enlarged and may feel tender to the touch.

All individuals suffering from bacterial endocarditis may not exhibit all of the foregoing symptoms and signs. An individual having an infection of the right side of the heart might well exhibit signs in the lungs, since they are supplied with blood by the right side of the heart. Conversely, an individual who has an infection of the left side of the heart, the aorta, or the mitral valve, will be more likely to have systemic symptoms—emboli in the skin and organs, kidney involvement, enlargement of the spleen, and aneurysms.

In almost all cases, the infection can be controlled by one more of the various antibiotic drugs. However, the dosages must be large and prolonged to insure that the drugs destroy the bacteria. The usual period for antibiotic administration is about one month. In most cases, blood tests will show that the bacteria are resistant to one or more types of antibiotic, so that the treatment may be even longer.

Kidney malfunction caused by bacterial endocarditis may be permanent, restricting the patient to reduced activity.

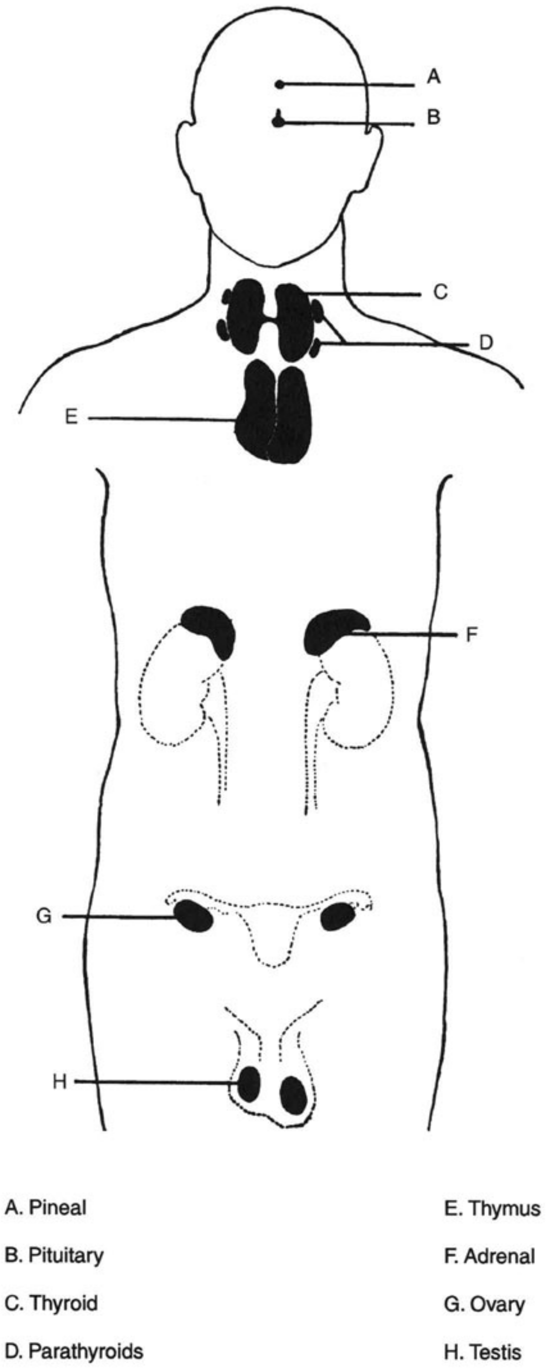
The individual with chronic heart disease should discuss this with the dentist or surgeon before undergoing tooth extraction, or simple ear, nose, or throat operations. These procedures may be especially dangerous, since bacteria from a throat infection or tooth abscess enter the blood stream in large numbers and, consequently, infect damaged areas of the heart.

ENDOCRINE SYSTEM. The endocrine system is made up of several organs situated in various parts of the body, all of which are characterized by their ability to produce active chemical substances called *hormones*. The organs of the endocrine system are called *glands of internal secretion* because they secrete the products of their activities directly into the blood stream for distribution throughout the body. *Endocrinology* is the science of the ductless glands and an *endocrinologist* is one who specializes in the science of the ductless glands and their function.

As indicated by the accompanying diagram, the endocrine glands are: (1) the *pituitary*, located at the base of the brain, which manufactures hormones that act as growth-stimulators of children, assist in the process of sexual maturation, and assist in the coordination of the activity of other endocrine glands, is thus sometimes referred to as the "master gland"—therefore, disorders of the pituitary can have grave significance; (2) the *hypothalamus*, which is part of the diencephalon, the central portion of the brain, and which is concerned with the regulation of many autonomic functions, including body temperature, sleep, behavior, appetite, and emotional response; (3) the *thyroid*, which is located in the neck, and which regulates the general speed of most of the chemical reactions in the cells of the body, regulating the calcium level in the blood, and playing a role in growth, development, and metabolism; (4) the *parathyroids*, located adjacent to the thyroid gland in the lower front portion of the neck and which maintain normal concentrations of calcium in many of the body tissues; (5) the *gonads (testes* in the male; *ovaries* in the female), which are the fundamental organs of reproduction and which produce gonadal hormones (respectively male and female hormones) that set secondary sex characteristics; (6) the *adrenals*, sometimes referred to in humans as the *suprarenal glands*, located near the top of the kidneys, and which secrete numerous hormones, notably *adrenalin* (epinephrine), which increases blood pressure, speeds up the respiration and the heartbeat, augments the amount of sugar in the blood, and in stressful situations, gives the individual a feeling of added strength and aggressiveness; (7) the *thymus*, which lies just above the heart, and which plays a central role in establishing the immunological capacities of the body, producing a hormone which is responsible for the production of cells with the capacity to make antibodies and reject foreign elements; (8) a part of the *pancreas*, notably the *islets of Langerhans* (see **Diabetes Mellitus**) which secrete insulin, glucagon, and gastrin, important in the regulation of sugar levels in the blood; and (9) sometimes classified as part of the endocrine system, the *pineal gland*, a small gland attached to the posterior part of the brain, behind and above the third ventricle, the functions of which remain somewhat obscure, but which secretes a hormone known as melatonin, an agent affecting pigment cells. Additionally, the *placenta* (the spongy structure in the uterus through which the fetus is nourished) also secretes hormones and thus is closely related to the endocrine system, and functional during pregnancy.

Some endocrine glands also perform other activities. For example, the pancreas produces a digestive secretion which is passed into the small intestine through a system of ducts.

Hormones fall into two general categories: (1) *messengers*, and (2) *managers*—to use a simple analogy. Most of the endocrine glands pro-



Distribution of the major endocrine glands in the human body.

duce one or more of each type. The messengers are those hormones which act upon tissues and organs outside of the endocrine system to speed or slow their normal functions. The managers also carry messages, but always within the endocrine system. They are responsible for the fact that some endocrine actions occur constantly, some occur periodically, some during certain years only, and some occur only once during the life of the individual.

One of the principal functions of the nervous system is to correlate all the parts of the body so that they may function harmoniously as a unit. In achieving this harmony, the nervous system is supplemented by the endocrine system; the two systems mutually affect each other. Mental strain, fear, anger, or any emotional state affects the activities of the endocrine glands. A well-known action is the accelerated production of adrenalin during a state of rage or fear. More subtle changes occur constantly in the endocrine system, elicited by some form of nervous stimulation. In some instances, the changes in endocrine activity are reflected in perceptible acts; in others, they remain imperceptible. The crying spells, hot flashes, and irritability sometimes seen in women during premenstrual tension or menopause illustrate this point.

The hypothalamus is part of both the nervous system and the endocrine system. The nerves connecting it to the pituitary as well as to the centers of neurosecretion permit a direction and indirect influence by the nervous system upon its endocrine counterpart. This influence is all the greater because the pituitary, in turn, has a definite influence over other endocrine glands. The hypothalamus also produces hormones which stimulate and which inhibit the pituitary gland. See **Nervous System and the Brain.**

See also **Addison's Disease; Adrenal Glands; Diabetes Insipidus; Gland; Gonads; Hormones; Parathyroid Glands; Pineal Gland; Pituitary Gland; Thymus Gland;** and **Thyroid Gland.**

For references see the lists at the ends of the aforementioned entries.

ENDODERM. See **Embryo.**

ENDOGENIC. A term used by geologists to denote processes originating within the earth.

ENDOGENOUS. Originating and developing from within—without influence of external factors and stimuli. In medical usage, a disorder, frequently of a biochemical nature, which arises "naturally" from and within a cell or organ in the absence of external contributing factors. Such malfunctions may arise from genetic abnormalities that are present with an individual entity and require no stimuli to generate symptoms of their presence with the exception of the passage of time.

ENDOMORPHISM. That phase of contact metamorphism which takes place in the intrusive magma rather than in the walls of the rock mass which it invades.

ENDOSCOPE. An instrument equipped with a lighting and lens system used for visual examination of the interior of a body organ or cavity.

ENDOSMOSIS. A type of osmosis in which the solvent dialyzes into the system. Exosmosis is the reverse process. The two processes may be illustrated by the conditions in the living cell; when the plasma is hypertonic, solvent passes from the cell into the plasma (exosmosis); when the plasma is hypotonic the solvent passes from the plasma into the cell (endosmosis).

The movement of the liquid relative to colloidal particles under an applied electrical field is termed electroendosmosis.

See also **Ocean Resources (Energy).**

ENDOTHELIUM. The delicate lining of the organs of circulation. It is one cell in thickness and is continuous throughout the closed passages with the exception of the sinusoids. The walls of capillaries are made up of little more than the endothelium.

END POINT (Chemical Reaction). See **Chemical Reaction Rate**

END POINT (Distillation). See **Petroleum.**

END-STAGE RENAL DISEASE (ESRD). See **Kidney and Urinary Tract.**

ENEMA. See **Constipation.**

ENERGY. In most contemporary texts and those of the last several decades, energy generally has been defined simply as "the ability or capacity to do work." This is a broadening of the earlier definition in terms of Newtonian mechanics which was "a property of moving masses."

The concept of energy is central to thermodynamics, quantitative chemistry, and electromagnetism. Consider Einstein's mass-energy equation, $E = mc^2$ for the interconversion of mass and energy, where E = energy in ergs; m = mass in grams; and c is the velocity of light in centimeters per second. Or, Planck's equation, which expresses the fundamental law of quantum theory, stating that the energy transfers associated with radiation are made up of definite quanta of energy proportional to the frequency of the radiation: $E = h\nu$, where E = the value of the quantum units of energy; ν = the frequency of radiation; and h is the elementary quantum of action, more commonly known as Planck's constant (6.6256×10^{-27} erg-second—the proportionality factor that, when multiplied by the frequency of a photon, gives the energy of the photon).

Although the fundamental definition of energy can be brief, it immediately calls for an explanation of work, and of power. In the strict physical sense, work is performed only when a force is exerted on a body while the body moves at the same time in such a way that the force has a component in the direction of motion. The amount of work done during motion from point "a" to point "b" can be expressed by

$$W = \int_a^b F \cos \theta \, ds$$

where F is the total force exerted and θ is the angle between the direction of F and the direction of the elemental displacement, ds. In the cgs system, the unit of work is the *dyne-centimeter* or *erg*; in the mks system, the *newton-meter* or *joule*; and in the English system, the *foot-pound.*

In rotational motion, the definition just given can be exactly applied, but it is often convenient to express the force as a torque and the motion as an angular displacement. The work done will be

$$W = \int_a^b \tau \cos \theta \, d\omega$$

where in this case, θ is always the angle between the torque τ, expressed as a vector quantity and the elemental angular motion $d\omega$, also expressed as a vector. The units of work performed in angular motion will, of course, be the same as in the case of linear motion. Notice that the definition of work involves no time element.

Power is defined as the rate at which work is performed. The average power accomplished by an agent during a given period of time is equal to the total work performed by the agent during the period, divided by the length of the time interval. The instantaneous power can be expressed simply as

$$P = dW/dt$$

In the cgs system, power has the units of *ergs per second*; in the mks system, units of *joules per second* (or *watts*); and in the English system, units of *foot-pounds per second.* A common engineering unit is the *horsepower*, defined as 550 foot-pounds per second; or 33,000 foot-pounds per minute. The SI unit of power is the *watt.* 1 watt = 1 joule per second. (1 joule is the work done by 1 newton acting through a distance of 1 meter.) 1 joule = 1 watt-second = 10^7 ergs = 10^7 dyne-centimeters. The SI unit of force is the newton. (1 newton = 10^5 dynes). See also entry on **Units and Standards.**

Now, returning to the basic definition of energy as the capacity for performing work. This definition may be better understood when stated as: "The energy is that which diminishes when work is done by an amount equal to the work so done." The units of energy are identical with the units of work previously given.

Energy can exist in a variety of forms, some more recognizable as being capable of performing work than others. Forms in which the energy is not dependent upon mechanical motion are generally referred to as forms of *potential energy*. The most common example in this category is gravitational potential energy. A body near the earth's surface undergoes a change in potential energy when it is changed in elevation, the amount being equal to the product of the weight of the body and the change in elevation.

Potential energy also may be stored in an elastic body, such as a spring or a container of compressed gas. It may exist in the form of chemical potential energy, as measured by the amount of energy made available when given substances react chemically. Potential energy also exists in the nuclei of atoms and can be released by certain nuclear rearrangements.

Kinetic energy is the energy associated with mechanical motion of bodies. It is quantitatively equal to $\frac{1}{2}mv^2$, where m is the mass of a body moving with velocity v. In the case of rotational motion, the kinetic energy is more easily calculated, using the expression $\frac{1}{2}I\omega^2$, where I is the moment of inertia of the body about its axis of rotation and ω is the angular velocity. Kinetic energy, like all forms of energy, is a scalar quantity (having magnitude but not direction). In a system made up of an assembly of particles, such as a given volume of gas, the total kinetic energy is equal to the sum of the kinetic energies of all the molecules contained in the volume. Calculation of the energy of such systems is very successfully treated theoretically on the basis of statistical averages.

Within a given system, energy may be transformed back and forth from one form to another, without changing the total energy of the system. A simple example is the pendulum, in which the energy is periodically converted from gravitational potential energy to kinetic energy and then back to gravitational potential energy. A similar situation, but on a submicroscopic scale, occurs in solid materials where the atoms are vibrating under the effect of interatomic rather than gravitational forces. As the temperature of a solid increases, the energy associated with the vibration of the atoms increases.

The example just given illustrates how, on a macroscopic scale, heat can be considered a form of energy. Regardless of the material involved, any amount of heat absorbed or released may be quantitatively expressed as an amount of energy. A *gram-calorie* of heat is equivalent to 4.19 joules, and in the English system, a *British thermal unit* (Btu) is equivalent to 778 foot-pounds.

Potential energy is also present in electric and magnetic fields. The energy available in a region of electric field is equal to $E^2/8\pi$ per unit volume, where E is the electric field strength. Within a given volume, the total energy represented by the electric field is the integral of $E^2/8\pi$ over the volume. Similarly, the energy represented by a magnetic field may be independently calculated by integrating $H^2/8\pi$ over any given volume, where H represents the magnetic field strength. In the case of an electrically charged capacitor, the total energy in the electric field, and hence in the capacitor, can be shown to be $\frac{1}{2}CV^2$. Here C is the capacitance and V the electric potential to which the capacitor is charged. Similarly the total energy in the magnetic field associated with an inductor carrying an electric current is $\frac{1}{2}LI^2$, where L is the inductance and I is the current.

Electromagnetic radiation is a combination of rapidly alternating electric and magnetic fields. Energy is associated with these fields and is exchanged between the electric and magnetic forms. This energy in a quantum of electromagnetic radiation, such as light or gamma radiation, can be expressed in different ways, but is commonly expressed as $E = h\nu$, as previously mentioned.

For particulate radiation or any very rapidly moving mass, the expression previously given for the kinetic energy, $\frac{1}{2}mv^2$, is not accurate when the velocity approaches that of the velocity of light. The theory of relativity requires a correction be made, and the exact kinetic energy, T, may be calculated in terms of the mass, m_0, of light in vacuum, c, as follows:

$$T = m_0c^2 \left[\left(1 - \frac{v^2}{c^2} \right)^{-1/2} - 1 \right]$$

Notice that this formula may also be written:

$$T = (m - m_0)c^2$$

where m is the variable quantity $m_0 \, (1 - (v^2/c^2))^{-1/2}$. This quantity represents the mass of the body, reducing to m_0 when v is zero, and approaching infinity as v approaches the speed of light.

This example illustrates another result of the theory of relativity, namely, the equivalence of mass and energy. Rewriting the last equation,

$$m = m_0 + \frac{T}{c^2}$$

The mass is seen to increase linearly with the kinetic energy of the body, the proportionality factor being c^2. It should be noted that even the rest mass, m_0, represents an amount of energy equal to m_0c^2. The total energy of a body of mass, m, can be generally given as:

$$E = mc^2 \quad \text{or} \quad E = m_0c^2 + T$$

In dealing with radiation, whether particulate or electromagnetic, it is customary to express energy in terms of electron volts. An electron volt is equal to the amount of work done when an electron moves through an electric field produced by a potential difference of one volt. One electron volt is equivalent to 1.60×10^{-12} erg. When charged particles, such as electrons or protons, are given kinetic energy by an accelerator, their kinetic energy is stated in terms of electron volts (eV), million electron volts (MeV), or billion electron volts (BeV).

A basic principle of physics known as the conservation of energy requires that within any closed system, the total energy must remain constant. Energy can be changed from one form to another; but the total, so long as no energy is added to or lost from the system, must be constant. In the case of the swinging pendulum, decreases in kinetic energy reappear as increases in potential energy and vice versa. Eventually, of course, the pendulum will stop due to the effect of frictional forces. At that time, all of the kinetic energy and gravitational potential energy will have been converted to heat.

In another example involving a radioactive atom, the total energy represented by the atom and the emitted radiation must be constant. If a gamma ray is emitted, the rest mass of the atom will be decreased by an amount equivalent to the sum of the energy of the gamma ray and the recoil kinetic energy of the atom, which will be very small. If a beta ray is emitted, the rest mass of the atom will be decreased by an amount equivalent to the sum of the rest mass of the emitted electron, the kinetic energy of the electron, and the recoil kinetic energy of the atom.

Entropy. In the mathematical treatment of thermodynamic processes there occurs very often a quantity, now relating energy to absolute temperature, now associated with the probability of a given distribution of momentum among molecules, and again expressing the degree in which the energy of a system has ceased to be *available energy*. Its mathematical form suggests that these are all aspects of a single physical magnitude. Application of the second law of thermodynamics leads to the conclusion that if any physical system is left to itself and allowed to distribute its energy in its own way, it always does so in a manner such that this quantity, called entropy, increases; while at the same time, the available energy of the system diminishes. This has led to the observation of a so-called "order of merit" for the various forms of energy.

Form of Energy	Entropy Per Unit Energy
Gravitation	0
Energy of rotation	0
Energy of orbital motion	0
Nuclear reactions	10^{-6}
Internal heat of stars	10^{-3}
Sunlight	1
Chemical reactions	1–10
Terrestrial waste heat	10–100
Cosmic microwave radiation	10^4

With reference to the foregoing listing, the energy usually flows from higher levels to lower levels—in a direction such that the entropy increases. Thus, cosmic microwave background radiation is defined as the ultimate heat sink, i.e., it represents the ultimate in energy degradation with no lower form in which to be converted.

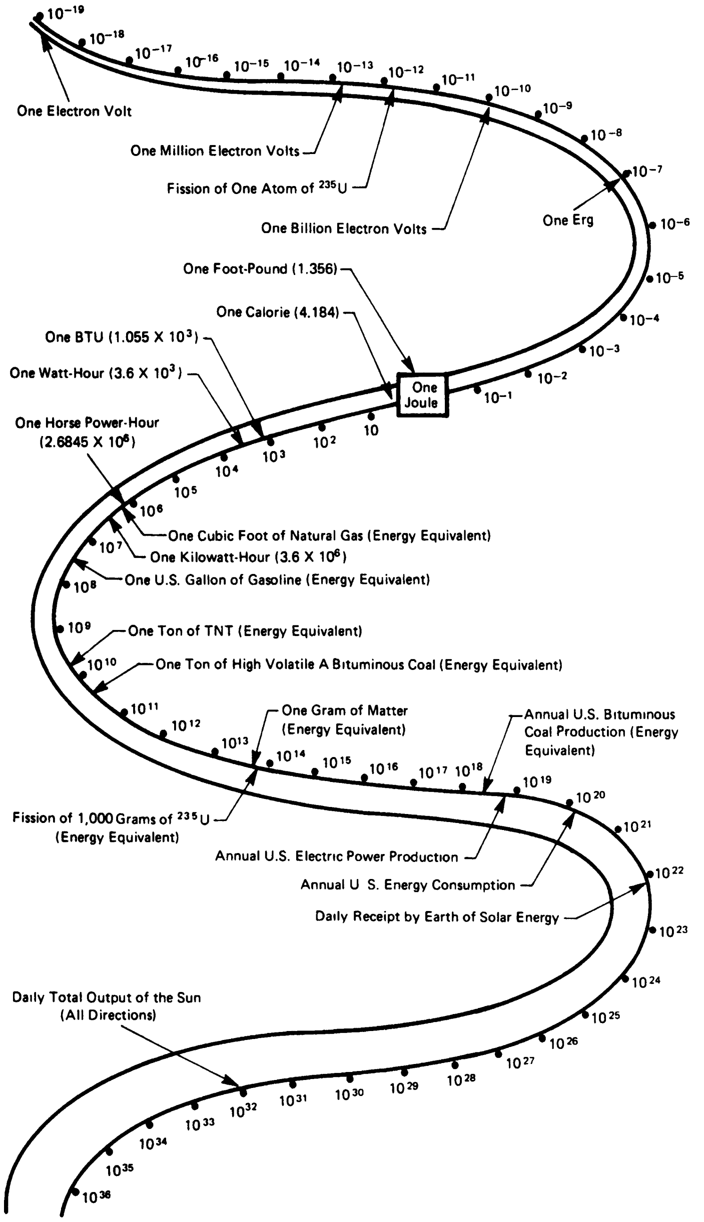
The universe evolved by the gravitational contraction of objects of all sizes, from clusters of galaxies to planets. In considering that thermodynamics appears to favor the degradation of gravitational energy to other forms, why is it that after an estimated 10 billion years since cosmic evolution, gravitational energy remains the predominant form? This is explained in terms of a series of phenomena which can be termed "hangups." These are, in essence: (1) The cosmos is large to the extreme; distances between objects are tremendously long; the average density is extremely low. Thus, matter cannot collapse gravitationally in a time shorter than the "free fall time." In relating free fall time (t) with density (d) with the formula, $Gdt^2 = 1$, where G is the constant in Newton's law of gravitation, it is apparent that the free fall time is extremely long. It is estimated that the time is about 100 billion years. But, since the density of our own galaxy is estimated at one million times that of the universe, more than the size hangup is required to preserve the galaxy. (2) An extended object cannot collapse gravitationally if it is spinning rapidly. The object assumes a stationary orbit revolving about the inner parts instead of collapsing. Thus, the earth has not collapsed into the sun. Other examples of the spin hangup at work include galaxies, planetary systems, double stars, and the rings of Saturn. (3) Hydrogen "burns" to form helium when it is heated and compressed. But, this thermonuclear burning releases energy, which opposes any further compression. Thus, a star with a lot of hydrogen cannot collapse gravitationally beyond a certain point until the hydrogen is burned up. It is estimated that the sun has been "stuck" on this thermonuclear hangup for 4.5 billion years and will need another 5 billion years to burn hydrogen before its gravitational contraction can be resumed. (4) Whereas a thermonuclear bomb is made mainly of heavy hydrogen, the sun contains ordinary hydrogen with only a trace of the heavy hydrogen isotopes. Whereas heavy hydrogen can burn explosively by strong nuclear reactions, ordinary hydrogen can react with itself only by the weak-interaction process. This proton-proton reaction proceeds about 10^{18} times more slowly than a strong nuclear reaction at same density and temperature. At least three fortunate circumstances contribute to the weak-interaction hangup: (a) without it, there would not have been a long-lived and stable sun; (b) the ocean would constitute an excellent thermonuclear high explosive; (c) hydrogen has survived rather than having been consumed in the initial, hot, dense phase of the evolution of the universe. (4) Because the transport of energy from the hot interior of the earth to the surface requires billions of years, the earth remains geologically active, these processes deriving their energy from the original gravitational condensation of the earth estimated as some 4 billion years ago. (5) There also is a special surface tension hangup, accounting for the survival of fissionable uranium and thorium nuclei in the earth's crust. They contain a high positive charge and excessive electrostatic energy such that they are ready to explode when triggered. However, before this can happen their surface must be stretched into a nonspherical shape. This process is opposed by an extremely powerful force of surface tension, estimated at about 10^{18} times stronger than that of a drop of water. Thus, it is estimated that fewer than one in a million of the earth's uranium nuclei fission spontaneously.

Energy Technology

Breadth in the "Packaging" of Energy. In the accompanying diagram, a packet of energy of 1 joule (1 newton-meter) is represented by the box in the upper center. Various energy packets, ranging from 1 electron volt (10^{-19} joule) to the daily energy output of the sun (total—in all directions) of 10^{32} joules are indicated.

Perfecting contemporary energy resources and power generation and consumption and developing presently nonconventional energy resources and systems possibly pose the greatest challenge to scientists and technologists in the last quarter of this century. In particular, scientists and technologists must be encouraged to work better together in an effective and realistic fashion to create constructive solutions to the problem of the energy/environment interface.

There are numerous entries in this encyclopedia on various energy topics. Consult the alphabetical index for such energy sources as coal, electric power, fuel cells, geothermal energy, hydroelectric power, hydrogen as a fuel, natural gas, nuclear power, oil shale, petroleum, solar energy, substitute natural gas and other synthetic fuels, tar sands, tidal energy, and waste materials. Also, a number of energy-converting and generating processes are described, including boilers and combustion, as well as energy-utilizing systems, such as diesel engines, gas and expansion turbines, internal combustion engines, steam engines, and steam turbines.



Spectrum of various energy quantities. (*Source: Omnibix U.S.A.*)

Additional Reading

Asbury, J. G., Geise, R. F., and R. O. Mueller: "Electric Heat," *Technology Review (MIT)*, **82**, 3, 32–46 (1980).

Burwell, C. C.: "The Role of Electricity in American Industry," Institute for Energy Analysis, Oak Ridge Associated Universities, Oak Ridge, Tennessee, June 1985.

Burwell, C. C., and D. L. Phung: "The Role of Electricity in Home Heating," Oak Ridge Associated Universities, Oak Ridge, Tennessee, February 1986.

Cleveland, C. J., et al.: "Energy and the U.S. Economy: A Biophysical Perspective," *Science*, **225**, 890–897 (1984).

Considine, D. M. (editor): "Energy Technology Handbook," McGraw-Hill, New York, 1976.

Devine, W. D., Jr.: "An Historical Perspective on the Role of Electricity in Manufacturing," Oak Ridge Associated Universities, Oak Ridge, Tennessee, January 1986.

Diennes, L., and T. Shabad: "The Soviet Energy System," Winston, Washington, D.C., 1979.

Dyson, F. J.: "Energy in the Universe," *Sci. Amer.*, **224** 3, 50–59 (1971).

Goeller, H. E., and A. Zucker: "Infinite Resources: The Ultimate Strategy," *Science*, **223**, 456–461 (1984).

Hayes, E. T.: Energy Resources Available to the United States, 1985–2000," *Science*, **203**, 233–239 (1979).

Hoogendoorn, C. J., and N. H. Afgan (editors): "Energy Conservation in Heating, Cooling, and Ventilating Buildings," 2 volumes, Hemisphere Publishing, Washington, D.C., 1978.

Hu, D. S.: "Handbook of Industrial Energy Conservation," Van Nostrand Reinhold, New York, 1982.

Kalhammer, F. R.: "Energy-Storage Systems," *Scientific American*, **241**, 6, 56–65 (1979).

Marlay, R. C.: "Trends in Industrial Use of Energy," *Science*, **226**, 1277–1283 (1984).

Paisson, B. O., et al.: "Biomass as a Source of Chemical Feedstocks," *Science*, **213**, 513–571 (1981).

Pound, R. V.: "Radiant Heat for Energy Conservation," *Science*, **208**, 494–495 (1980).

Reynolds, J. Z.: "Power Plant Cooling Systems: Policy Alternatives," *Science*, **207**, 367–372 (1980).

Staff: "Energy in Transition—1975–2010," Freeman, San Francisco, 1980.

Staff: "World List of Nuclear Power Plants," *Nuclear News*, **29**(10), 77–96 (August 1986).

Staff: "Two Energy Futures: National Choices Today for the 1990s," American Petroleum Institute, Washington, D.C., 1986.

Staff: "1986 Research & Development Program Plan," Electric Power Research Institute, Palo Alto, California, 1986.

Staff: "U.S. Energy Outlook—Fuels for Electricity," National Petroleum Council, Washington, D.C. (Revised periodically).

Staff: "Statistical Year Book on the Electric Utility Industry," Edison Electric Institute, New York (Published annually).

Staff: "Semi-Annual Electric Power Survey," Edison Electric Institute, New York (Published semi-annually).

Tybach, L., and L. J. P. Miffler (editors): "Geothermal Systems," Wiley-Interscience, New York, 1981.

Ward, G. M., Sutherland, T. M., and J. M. Sutherland: "Animals as an Energy Source in Third World Agriculture," *Science*, **208**, 570–574 (1980).

Zaleski, C. P. L.: "Energy Choices for the Next 15 Years: A View from Europe," *Science*, **203**, 849–851 (1979).

ENERGY BALANCE (Planet). See **Heat Balance (Planet).**

ENERGY CONSERVATION (Conservation Laws and Symmetry. See **Insulation (Thermal).**

ENERGY (Fuel Cell). See **Fuel Cells.**

ENERGY (Geothermal). See **Geothermal Energy**

ENERGY (Kinetic). See **Kinetic Energy.**

ENERGY LEVEL. A stationary state of energy of any physical system. The existence of many stable, or quasi-stable, states, in which the energy of the system stays constant for some reasonable length of time, is an essential characteristic of quantum-mechanical systems, and is the basis of large areas of modern physics.

ENERGY LOSS (Combustion). See **Combustion.**

ENERGY (Nuclear). See **Nuclear Power Technology.**

ENERGY (Potential). See **Potential Energy.**

ENERGY (Solar). See **Solar Energy.**

ENERGY STATE TERMS. Terms designating the discrete energy states of a particle in a system. Thus the energy states of an atom are called $S, P, D, F, \ldots$ terms, respectively, corresponding to the values 0, 1, 2, 3, . . . of L, the resultant angular momentum quantum number of the atom. The energy states of a molecule are called $\mathbf{\Sigma}, \mathbf{\Pi}, \mathbf{\Delta}, \mathbf{\Phi} \ldots$ terms, respectively, corresponding to the values 0, 2, 3, . . . of λ, the electronic orbital angular momentum (about internuclear axis) quantum number.

The letters indicating the value of L are usually preceded by a superscript denoting the multiplicity and followed by a subscript denoting the total angular momentum quantum number J. In addition, the principal quantum number is often written as a coefficient. Energy state terms and their transitions are shown in energy level diagrams.

Magnetic Energy State. A magnetic dipole of a moment μ in a magnetic field of flux density B has an energy that depends on orientation, $E = -\mu B \cos\theta$, or in vector notation $E = -\boldsymbol{\mu}\cdot\mathbf{B}$, in mksa units ($-\mu\cdot\mathbf{H}$ in emu). In atomic and nuclear systems the orientation of $\boldsymbol{\mu}$ relative to $\mathbf{B}$ is quantized, only certain values of $\cos\theta$ being allowed. Transitions between these allowed magnetic energy states may take place with the emission or absorption of electromagnetic (magnetic dipole) radiation of frequency given by the Bohr condition:

$$\nu = \Delta E/h, \qquad \text{or} \qquad \omega = \Delta E/\hbar$$

Particles such as electrons, protons, nuclei, etc., have intrinsic magnetic moments $\boldsymbol{\mu} = eg\hbar\mathbf{I}/2M$, where $\mathbf{I}\hbar$ is the spin angular momentum, g the g-factor of the particle, and M is the mass of the electron or the mass of the proton. The magnetic energy states are thus given by $E = -\hbar eg\mathbf{I}\cdot\mathbf{B}/2M = -\hbar egmB/2M$. Here m is the magnetic quantum number, which can take on the values $-I, -(I-1), \ldots (I-1), I$ where I is the spin quantum number ($\frac{1}{2}$ for electrons and protons). The energy is also often written $E = -\mu_B gmB$ or $-\mu_N gmB$ where μ_B and μ_N are the Bohr magneton and nuclear magneton respectively. The magnetic quantum number can only change by ± 1 as a result of the emission of radiation, so that there is only one emission or absorption frequency $\omega = egB/2M$.

Negative Energy State. 1. Any bound state, in which the sum of the kinetic energy and the potential energy, the latter reckoned relative to zero at infinity, is less than zero. The existence of such states is essential for the stability of any system that is not surrounded by a region of positive potential energy, such as the Coulomb barrier.

2. A consequence of the Dirac electron theory is that there exist electron states of *negative* total energy (including both rest mass energy and kinetic energy). Electrons in such states of negative energy are unobservable, only electrons of positive total energy being observable. The allowed positive and negative states are shown in the diagram (only $E > m_0c^2$ and $E < -m_0c^2$ are allowed in a field-free region). If a γ-ray photon of energy greater than $2m_0c^2$ (where m_0 is the rest mass energy of the electron) is absorbed by an electron of negative energy, it will be lifted into a positive energy state and will become observable. The positron is identified with the hole that is left behind.

ENERGY (Tidal). See **Tidal Energy.**

ENERGY UNITS. See **Units and Standards.**

ENGINE. In common usage, the term engine is used widely for devices which produce motion. In stricter technical sense, an engine is said to transform energy, especially heat energy, into mechanical work. Among the prime movers, those in which the power originates in a piston and cylinder are classed as engines, while those with purely rotative motion are known as turbines.

ENGINE (Gas). See **Gas and Expansion Turbines.**

ENGINE (Four-Cycle). See **Four-Cycle Engine.**

ENGINE (Internal Combustion). See **Automotive Electronics; Internal Combustion Engine.**

ENRICHMENT. 1. Also "secondary enrichment." The term applied by students of ore deposits to the natural processes by which the lower levels of an ore deposit are enriched at the expense of the upper levels, or the original protore. Particularly applied to lodes in which the sulfide ores have been concentrated by the leaching of the upper levels of the vein and redeposition below the groundwater table. Important ore minerals belonging to this type are chalcocite and argentite.

2. Any process which changes the isotopic ratio; in reference to uranium, it is a process which increases the ratio of ^{235}U to ^{238}U in uranium by separation of isotopes.

ENSEMBLE. A collection of similar systems considered in statistical mechanics. Ensembles were introduced by Gibbs, and their importance lies in the fact that the average behavior of a system in an ensemble can often be used to predict the behavior of an actual physical system. Usually, all systems in an ensemble are supposed to have the same number of constituent particles. Such ensembles are called *petit ensembles*. Examples of petit ensembles which are used extensively are the *microcanonical* and *macrocanonical ensembles*. In a *microcanonical ensemble*, the variation in energy (or other independent variable) of all the systems lies within an infinitesimal range. Over this range, the assembly is in statistical equilibrium. In a *macrocanonical ensemble*, there is present a collection of identical microcanonical ensembles. In both of them, and any other *canonical ensemble*, the distribution in energy of the system is given by the Boltzmann factor. In a *cooperative ensemble*, the interactions between the systems composing the ensemble are not negligible. The state of a given system is largely determined by the states of the neighboring systems, while in an *ideal ensemble*, such as a perfect gas or ideal solution, these interactions can be neglected.

The *density of an ensemble* is written as the quantity ρ defined in such a way that $\rho\, d\Omega$ is the fraction of systems in an ensemble which have values of the momenta and position coordinates of all the particles in the system corresponding to a point in gamma-space within the extension in phase $d\Omega$. For grand ensembles one must suitably alter this definition, *grand ensembles* being defined by Gibbs as ensembles which do not have the same numbers of particles. The most often used grand ensemble is the *grand canonical ensemble*, which has a density ρ defined by

$$\rho = e^{-q+\nu n-\beta\epsilon}$$

where q is a (normalizing) constant, $\beta = 1/kT$ (k, Boltzmann's constant; T, absolute temperature), ϵ, the energy of the system, $\nu = \beta g$ (g, the partial thermal potential), and n the number of particles of the system.

ENSIGN FLY (*Insecta*, *Hymenoptera*). Small parasitic insects whose abdomen is elevated on a slender stalk above the thorax. It has been likened to a flag and gives the common name to the group. The ensign flies make up the family *Evaniidae*. In all species whose habits are known, the larvae are parasitic in the eggs of cockroaches.

Ensign fly.

ENSTATITE. The mineral enstatite is an orthorhombic pyroxene, rarely in distinct crystals, usually found as fibrous or lamellar masses or perhaps compact. It has one easy cleavage parallel to the prism; brittle with uneven fracture; hardness 5–6; specific gravity 3.2–3.4; luster pearly to vitreous, sometimes somewhat metallic in bronzite, a variety of enstatite carrying up to 15% ferrous oxide, FeO. Color grayish to greenish or yellowish-white, green and brown. Chemically, enstatite is a silicate of magnesium, $MgSiO_3$. It occurs in igneous rocks which are high in magnesium content, like gabbros, diorites, and pyroxenites, and less commonly in metamorphic rocks. Meteorites of both the stony and metallic types have been shown to contain enstatite. It has been found at many places in Europe, (the former Czechoslovakia, Austria, Bavaria, Germany, Norway), and the Republic of South Africa. In the United States it occurs in Putnam and St. Lawrence Counties, New York; Lancaster County, Pennsylvania; Jackson County, North Carolina, and near Baltimore, Maryland. The name enstatite is derived from the Greek word meaning *opponent*, in reference to its refractory nature; it is almost infusible. See also **Pyroxene.**

ENTERIC CAVITY. The digestive cavity, *enteron*. This cavity forms by the splitting or invagination of the inner germ layer early in embryonic (see **Embryo**) development and persists as a sac with one opening to the exterior in the coelenterates and flatworms. In this form it is also called the archenteron.

In animals with a tubular alimentary tract the enteric cavity becomes the primitive gut. Its endodermal lining becomes the glandular digestive tissue and gives rise to large glandular masses in some species, and in the terrestrial vertebrates also produces the respiratory system.

ENTERITIS. Any inflammation of the small intestine, usually accompanied by fever, pain in the abdomen, diarrhea and other constitutional symptoms. The condition is commonly called gastroenteritis. See **Diarrhea.**

ENTHALPY. The *enthalpy*, H or *heat content*, of a substance is a thermodynamic property defined as the *internal energy*, E, plus the product of the pressure, P, times the *volume*, V, of the substance

$$\mathrm{H} = E + PV \tag{1}$$

The enthalpy is an extensive state function; its value depends only on the state and the amount of the substance and not on its previous history. It has the units of energy and it is usually expressed in calories (or kilocalories).

For a process at *constant pressure* ($\Delta P = 0$), in which the only work performed is the mechanical pressure-volume work ($P\,\Delta V$), the *change in enthalpy*, ΔH, is equal to the heat adsorbed by the system, q (hence the name heat content):

$$\mathrm{D}H = \Delta E + P\,\Delta V = q \tag{2}$$

This relation is a direct consequence of the definition of enthalpy by Equation (1) and of the mathematical statement of the first law of thermodynamics, namely that the change in internal energy, ΔE, is equal to the heat adsorbed minus the work done ($q - P\,\Delta V$). It is clear that this thermodynamic relation does not define absolute values of enthalpy or internal energy. Changes in enthalpy, however, are readily measured by calorimetric techniques, and the relative enthalpy values are sufficient for all thermochemical calculations.

Enthalpy-Temperature Relation and Heat Capacity. When heat is adsorbed by a substance, under conditions such that no chemical reaction or state transition occur and only pressure-volume work is done, the temperature, T, rises and the ratio of the heat adsorbed, over the differential temperature increase, is by definition the heat capacity. For a process at constant pressure (following Equation (2)), this ratio is equal to the partial derivative of the enthalpy, and it is called the *heat capacity at constant pressure*, C_p, (usually in calories/degree-mole):

$$\left(\frac{\partial H}{\partial T}\right)_p = C_p \tag{3}$$

The temperature dependence of H for a substance remaining in the same physical state can be expressed as a function of C_p by integration of Equation 3.

If a substance undergoes a transformation from one physical state to another, such as a polymorphic transition, the fusion or sublimation of a solid, or the vaporization of a liquid, the heat adsorbed by the substance during the transformation is defined as the *latent heat of transformation* (transition, fusion, sublimation or vaporization). It is equal to the enthalpy change of the process, which is the difference between the enthalpy of the substance in the two states at the temperature of the transformation. For the purpose of thermochemical calculations, it is usually reported as a molar quantity with the units of calories (or kilocalories) per mole (or gram formula weight). The symbol L or ΔH, with a subscript t, f (or m), s, and v is commonly used and the value is usually given at the equilibrium temperature of the transformation under atmospheric pressure, or at 25°C. For a substance undergoing one phase transformation, with a latent heat Δh_t, at a temperature T_t, the enthalpy change between two temperatures, T_1 and T_2, such that $T_1 < T_t < T_2$, is given by

$$H_T - H_T = \int_{T_1}^{T_t} C_p'\,dT + \Delta h_t + \int_{T_t}^{T_2} C_p''\,dT \tag{4}$$

where C_p' and C_p'' are the heat capacities of the substance in the two different physical states. For several successive transformations, additional terms are added. The accompanying diagram illustrates the temperature dependence of enthalpy and heat capacity.

Very precise measurements of the heat capacity of liquids and solids can be obtained by calorimetric techniques at relatively low temperatures (below 200°C) and they can be extrapolated down to the absolute zero of temperature (−273.15°C) by reliable theoretical expressions. In that temperature range, heat capacity data are usually very accurate and enthalpy values are obtained by integration (Equation (4)). The most reliable method for determining high-temperature enthalpies and heat capacities is the dropping method (or method of mixtures) which consists of dropping the substance under investigation from a furnace at a known temperature into a calorimeter at room temperature. This method determines directly the change in enthalpy (or heat content) of the substance between the temperature of the furnace and that of the calorimeter. Heat capacities are obtained by differentiation (Equation (3)). The measurement of heat capacity of gases is usually more difficult, and their thermodynamic properties can be more accurately calculated by methods of statistical mechanics based upon energy level of gas molecules obtained from spectroscopic data, or upon the knowledge of the molecular configuration and the vibration frequencies of the molecules.

Molar enthalpy data for elements and inorganic compounds above room temperature are usually tabulated in the form of the heat content above a reference temperature, usually 298.15°K = 25°C. They are represented by: $\mathrm{H}_T - H_{298.15}$ in calories/mole. The data are correlated over a range of temperature by empirical equations such as a series of powers of the absolute temperature or such as the following expression adopted by K. K. Kelley (1960) for his extensive compilation of data on inorganic compounds:

$$H_T - H_{298.15} = aT + bT^2 + c/T + d \tag{5}$$

where T is the absolute temperature (°K) and a, b, c, d are constants determined from experimental data. The corresponding equation for heat capacity is:

$$C_p = a + 2bT - c/T^2 \tag{6}$$

Standard Enthalpy of Formation. For the convenience of tabulation and computation of thermodynamic data, it is essential to present them in a commonly accepted form relative to a single standard state of reference. At all temperatures, the *standard state* for a *pure liquid or solid* is the *condensed phase under a pressure of 1 atmosphere*. The standard state for a *gas* is the *hypothetical ideal gas at unit fugacity* (equivalent to a "perfect gas" state), in which state the enthalpy is that of the real gas at the same temperature when the pressure approaches zero. Values of thermodynamic quantities for standard-state conditions are identified by a superscript ^{0}, and H^0, for instance, is the enthalpy change of a reaction when reactants and products are in the standard state.

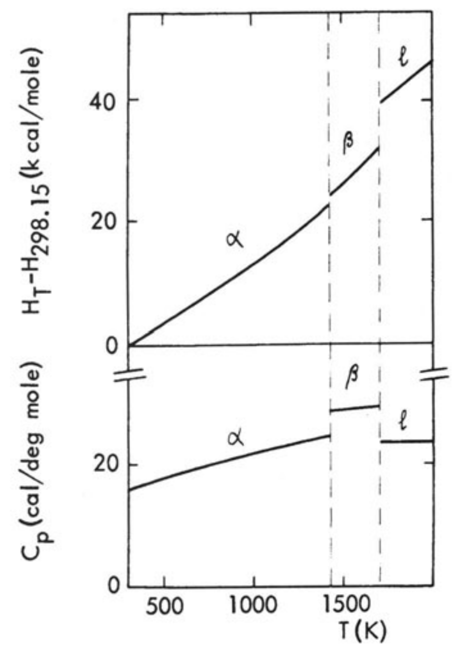
The *standard enthalpy of formation*, ΔHf^0 (also represented by ΔH_f^0) or simply H_f^0), of a substance at a given temperature is by definition, the enthalpy change when 1 mole of the substance in its standard state is formed, isothermally, at the indicated temperature from the elements, each in its standard state. Usual units are kilocalories/mole. *For all elements* in their *stable form at 25°C* (298.15°K), the *enthalpy of formation is zero*. If solid substances have more than one crystalline form, the most stable one is taken as the standard state, and the others have slightly different enthalpies. This convention about zero enthalpy is arbitrary but universally accepted, and it may be compared to the arbitrary choice of zero for terrestrial altitudes. The combination of enthalpies of formation, enthalpies of transition, and heat capacities makes possible the calculation of the enthalpy of a substance, in a given state at a given temperature, relative to a commonly accepted reference.

Enthalpy calculation for mixtures is more complex than for pure substances and its discussion is beyond the scope of the present article. *Aqueous solutions* (q.v.), however, are very important from a geochemical point of view, and reliable data are usually available. The *enthalpy of solution* is the enthalpy change resulting from the dissolution of a substance; it is a function of the solute concentration and its values are reported accordingly in the literature. For a *solute in aqueous solution*, a *standard state* is defined as the *hypothetical ideal solution of unit molality* (1 mole of solute per 1000 grams of water). In this state, the partial molal enthalpy and heat capacity of the solute are the same as in the infinitely dilute real solution. Although it is impossible to prepare a solution of only one ionic species (since the system must remain neutral), it is convenient to apportion the enthalpy (and other thermodynamic properties) between the various ions. This appointment is not unique and an additional convention has to be made, namely that the *standard enthalpy of hydrogen* ion in aqueous solution (aq) at unit activity, ΔHf^0 for H^+ (aq), *is zero*. The properties of a neutral electrolyte in the standard state are equal to the algebraic sum of the values corresponding to the individual ions.

Heat of Reaction and Gibbs Free Energy Change. For a chemical reaction, at constant pressure with only pressure-volume work performed, the heat adsorbed by the process, q, is equal to the enthalpy change, ΔH, or the sum of enthalpies of the products of the reaction minus the sum of enthalpies of the reactants (taking into account the amount of each).

$$q = \Delta H = \Sigma H_{\text{products}} - \Sigma H_{\text{reactants}} \tag{7}$$

Those heat effects can be easily calculated when the enthalpies of formation and the enthalpy-temperature relations are available for the substances considered. Usually, the *heat of reaction* is defined as the heat evolved by the process, and it is equal to the enthalpy change but opposite in sign, while heats of fusion or vaporization always refer to the heat adsorbed, and for heats of solution the usage varies. In order to avoid any confusion, it is recommended to express heat effects of chemical process by reporting the enthalpy change, ΔH.



Example of the temperature dependence of enthalpy relative to 25°C, $H_T - H_{298\ 15}$, and heat capacity, C_p. The data are for fluorite, CaF_2 (*K. K. Kelly, 1960*). The discontinuities in the lines correspond to the α to β transition (1424 K) and the fusion (1691 K).

Early chemists thought that the heat of reaction, $-\Delta H$, should be a measure of the "*chemical affinity*" of a reaction. With the introduction of the concept of *netropy* (q.v.) and the application of the second law of thermodynamics to chemical equilibria, it is easily shown that the true measure of chemical affinity and the driving force for a reaction occurring at constant temperature and pressure is $-\Delta G$, where ΔG represents the change in thermodynamic state function, G, called *Gibbs free energy* or *free enthalpy*, and defined as the enthalpy, H, minus the entropy, S, times the temperature, T ($G = H - TS$). For a chemical reaction at constant pressure and temperature:

$$\mathrm{D}G = \Delta H - T\,\Delta S \tag{8}$$

and the Gibbs free energy change can be obtained by calculating the enthalpy and entropy change and applying Equation (8). The criterion for *a spontaneous chemical reaction* is that ΔG be *negative*, and a chemical equilibrium corresponds to the condition $\Delta G = 0$.

Conversely, if the Gibbs free energy change is known as a function of temperature at constant pressure, the enthalpy change can be obtained by a relation which is an alternate form of the *Gibbs-Helmholtz equation*, and which can be derived from Equation (8).

$$\left(\frac{\partial(\Delta g/T}{\partial(1/T)}\right)_P \ \Delta H \tag{9}$$

This means that ΔH is the slope of the line representing $\Delta G/T$ versus $1/T$ at constant pressure.

For references, see entries on **Heat Transfer;** and **Thermodynamics.**

ENTOMOLOGY. The science that deals with all facts pertaining to insects. Because of the large number of insect species and their frequent economic importance, the principal divisions of the science have been systematic entomology and economic entomology. Classification is difficult and intricate, and demands the constant service of specialists. The economic field is of such importance, especially to agriculture, that the national and state governments maintain organizations for the scientific study of insects which also assist in their control. See **Insecticide and Insecticide Technology.**

Although entomologists may specialize in insect morphology or physiology, work of this type is less extensive than more practical studies and is of more general biological interest; hence, it takes its place largely in the subsidiary sciences of general biology.

ENTRANCE SLIT. A narrow slit in an opaque screen through which light enters a spectrometer. The spectrum formed is the image of this slit in each wavelength of light present. A narrow slit is necessary for good resolution to avoid great overlapping of these images. However, the smaller the entrance slit, the less radiation enters the spectrometer. Hence a slit width must be used which is a compromise between the resolution desired and the necessary light intensity for proper observation or detection.

ENTRENCHED MEANDER. Also termed *incised meander*, a river valley which has a distinctly meandering old-age pattern (longitudinal profile), and a V-shaped or canyon-shaped, youthful transverse profile. The meandering course of the river is inherited from the time when it flowed at, or close to, base level, that is on a relatively flat surface. Subsequent uplift of the region quickens the flow of the stream, hence its downcutting power, without necessarily altering its inherited meandering course. Entrenched meanders are therefore physiographic evidence of the rejuvenation of the erosive power of an old-age stream. Entrenched meanders may suggest the first stage in a new cycle of erosion, but usually imply an interruption in the normal cycle due to relatively sudden uplift of the region before the entire area has been reduced to a peneplain.

A rejuvenated region showing entrenched meanders. (Yakima Canyon, Washington.)

ENTROPY. 1. In the mathematical treatment of thermodynamic processes there occurs very often a quantity, now relating energy to absolute temperature, now associated with the probability of a given distribution of momentum among molecules, and again expressing the degree in which the energy of a system has ceased to be available energy. Its mathematical form suggests that these are all aspects of a single physical magnitude. Application of the second law of thermodynamics leads to the conclusion that if any physical system is left to itself and allowed to distribute its energy in its own way, it always does so in a manner such that this quantity, called "entropy," increases; while at the same time the available energy of the system diminishes. This law applies to the universe as a whole, hence the proposition that the total entropy increases as time goes on. An interesting conclusion as to entropy in the vicinity of absolute zero is expressed by the Nernst heat theorem; viz, that all physical and chemical changes in this region take place at constant entropy. Any process during which there is no change of entropy is said to be "isentropic." This is true, for example, of an adiabatic process in which there is no dissipation of energy, i.e., one which is also a reversible process. In thermodynamics discussions entropy is commonly classed, along with temperature, pressure, and volume, as one of the variables defining the state of a body, and is often graphed as such on thermodynamic diagrams.

2. In information theory, entropy is a measure of the uncertainty of our knowledge.

3. In thermodynamics, entropy is defined by the equation

$$dS = dQ/T$$

where dS is an infinitesimal change in the entropy of a system, dQ is the infinitesimal amount of heat that enters the system, and T is the absolute temperature.

In statistical mechanics, entropy is

$$k \log_e P + \text{constant}$$

where k is Boltzmann's constant, and P is the statistical probability of the state considered.

Standard Entropy. The total entropy of a substance in a state defined as standard. Thus, the standard states of a solid or a liquid are regarded as those of the pure solid or the pure liquid, respectively, and at a stated temperature. The standard state of a gas is at 1 atmosphere pressure and specified temperature, and its standard entropy is the change of entropy accompanying its expansion to zero pressure, or its compression from zero pressure to 1 atmosphere. The standard entropy of an ion is defined in a solution of unit activity, by assuming that the standard entropy of the hydrogen ion is zero.

Entropy of Disorder. That part of the entropy of a substance that is due to a disordered arrangement of the particles as opposed to a similar but ordered arrangement. The most clear-cut example is the order-disorder transition in binary alloys, in which virtually the whole entropy change is of this kind. The entropy change on fusion of a solid is largely due to entropy of disorder.

See also **Energy.**

ENTRY CORRIDOR. Depth of the region between two trajectories which define the design limits of a space vehicle which will enter a planetary atmosphere.

ENVELOPE (Mathematics). The equation of a curve usually contains one or more constants, in addition to the dependent and independent variables. If the constants are regarded as variable parameters, a family of curves is generated as these parameters are assigned a series of different value. In case only one such parameter is involved, the equation for the family of curves can be written as $f(x, y, t) = 0$. If another curve, or group of curves, is a common tangent to the given family, it is said to be the envelope and it can be found by solving the two equations.

$$f(x, y, t) = 0; \partial f/\partial t = f_t(x, y, t) = 0$$

to obtain x and y as a function of the parameter t.

The procedure can be generalized for families of curves which depend on two or more parameters.

See also **Circular Curves; Robot and Robotics;** and **Tangent** (Geometry).

ENVIRONMENT. The assemblage of material factors and conditions surrounding the living organism and its component parts.

Environment includes both external and internal factors. In the external environment inanimate objects and the forces associated with them constitute the physical environment, and the living things and their derivatices with which the animal may be associated constitute the organic environment. Within its body it maintains an organization which constitutes an internal environment to which all of its parts respond directly, whether or not they also have external contacts.

During the past few decades, the word *environment* has become a "household word" for numerous scenarios, but is used mainly to identify air, land, or water pollution. But this all-encompassing word also may refer to many more specific situations, such as to the deleterious effects of noise, nuclear and ionizing radiation, or electromagnetic radiation that emits from cathode-ray tubes and power lines; to the contributions of modern packaging to waste disposal and accompanying pollution; to the trade-offs between environmental and economic factors in selecting energy sources; or to the threat that light pollution may make obsolete some existing astronomical observatories—and the list goes on seemingly ad infinitum.

Thus, the staff of this encyclopedia elected to place these various environmental concerns in their proper places, distributed throughout the two volumes, coupled with a convenient list of environmentally related topics in the alphabetical index.

ENVIRONMENT (Controlled). Frequently, in industry and research facilities, it is necessary to provide a planned and carefully controlled environment in which to conduct manufacturing, assembly, inspection, and test operations as well as scientific investigations. Clean rooms, temperature-controlled rooms, sound rooms, and dry rooms are terms that are often applied to special types of controlled environments. Some environmental systems are designed to control a single condition; in other cases, many variables are controlled. A clean room may require only dust particle control and a comfortable temperature for the personnel who work in the room. In the case of a room for calibrating dimensional standards and for measuring precision parts, very close temperature and humidity also will be required. Controlled environments, of course, also are required in hospitals and medical facilities, in horticultural establishments, in museums, and numerous other industrial and commercial structures. There is a fine division between what may be termed a controlled environment and what may be inferred from the term *air conditioning*. Essentially a controlled environment differs from an air-conditioned space on two counts: (1) a controlled environment may include the measurement and very careful control of many environmental factors other than temperature and humidity which are the primary concerns of air conditioning, and (2) a controlled environment implies much greater precision in control and usually for reasons other than the comfort and well being of the occupants of a space. Further, a controlled environment may not even involve air, but rather may be concerned with the control of an aqueous environment, or of nonatmospheric factors, such as protection from shock, vibration, change of position, and so on.

Space does not permit a delineation of the many types of controlled environments encountered in industry and research. Because of the tight control requirements, an example is included of environmental control for a metrology standards laboratory and for so-called "clean rooms." The semiconductor industry also is a major user of clean rooms.

The International Organization for Standardization adopted 20°C (68°F) as the standard reference temperature for length measurement in 1951. Environmental systems for linear measurements provide temperature control at 20°C ± 0.03°C, maintain humidity between 40 and 45% (R.H.), and remove 99.97% of dust particles above 0.3 micrometer in size at the filter. Outside of the metrology laboratory, but in dimensional inspection rooms concerned with production parts, the temperature is controlled at 20°C ± 0.60°C.

Environmental rooms usually have air lock entry chambers. An air shower is included where critical control of duct contamination is required. A blast of air for a measured length of time is designed to remove foreign particles from an individual's outer clothing, usually a smock or uniform.

Generally, clean rooms using either vertical or horizontal laminar air flow are designed to control contamination and other variables as follows:

Particle Count—not to exceed 100 particles per cubic foot (approximately 3500 particles per cubic meter) of air of a size of 0.5 micrometer and larger.

Temperature—72°F ± 1°F (22.2°C ± 0.6°C) under static conditions in a horizontal plane 1 foot (0.3 meter) above the floor.

Relative Humidity—50% maximum. This depends upon use.

Air Velocity—60 feet (18 meters)/minute (minimum) to 100 feet (30 meters)/minute (maximum), down to the floor in a straight-line flow.

Make-up Air—15% minimum of total air used for air conditioning.

Positive Static Pressure—0.05 to 0.10 inch (1.3 to 2.5 millimeters) of water.

ENZOOTIC. Term describing any disease of animals whose incidence and distribution resemble those of endemic disease in man.

ENZYME. An enzyme is a protein that serves as a catalyst for a particular biological transformation—as, for example, the conversion of sugar into alcohol and water. Because of the make-up of the genetic material, most enzymes are highly specific. As discussed in the article on **Industrial Biotechnology,** this specificity is very advantageous in bioprocessing.

Fermentation, one of the most common transformations to be accomplished with the aid of an enzyme as catalyst, has been known for about 4000 years, mainly in connection with brewing, winemaking, and dairy products, such as cheese and yogurt. It was not until the early 1600s that the concept of an enzyme was recognized. Well over 300 years passed by, however, before the first enzyme to be isolated, *urease*, was produced in crystalline form. Shortly thereafter, numerous other enzymes were isolated in pure form, including amylase, carboxy-peptidase, chymopapain, papain, pepsin, and starch phosphorylase. Today, there are many hundreds of known enzymes with many specific purposes. Enzymes, once created from microorganisms right in the fermenting vessel, can in some instances be purchased in pure form for addition to the fermentation batch vessel. As mentioned in the article on **Enzyme Preparations,** purified enzymes are widely used in the food industry for enhancing flavor and stabilization of food quality and, among other uses, are compounded in packaged detergents. Major classes of enzymes include the oxidoreductases, transferases, hydrolases, lyases, isomerases, and ligases or synthetases, their names indicative of their functions.

Sources of Enzymes. Enzyme complexes are generated by living cells, notably yeasts, molds, bacteria, and actinomycetes. Enzymes are involved in numerous biological transformations, as in the metabolism of living organisms, and thus play a vital role at practically all levels of food involvement—production, processing, and consumption, whether by fish, bird, insect, or primate. Enzymes and the transformations which they promote are ever present during the entirety of the food chain. Investigations in botany, pursuits of agronomy, studies of nutrition, inquiries into plant and animal pathology, and the numerous other aspects of science that are involved in life processes, when probed in depth, ultimately encounter the vital roles played by enzymes.

Characteristics of Enzymes

Although with the advent of gene recombination technology the knowledge of enzymes is gaining rapidly, their principal characteristics have been established for decades. These include sensitivity to the environment (temperature and pH), the need for a clean watery medium (solvent), the requirements for nutrients, such as carbon (for energy), oxygen (absence or presence depending upon whether microorganism involved is anaerobic or aerobic), nitrogen, phosphorus, and trace substances.

Common properties of enzymes include:

1. Their predominant, established role as catalysts, often providing the means of effecting chemical (biological) conversions that otherwise would be difficult and at lower rates of energy expenditure.
2. Their structure which suggests that enzymes are simple or conjugated proteins.
3. Their relatively high sensitivity to environmental conditions.
4. Their origin from living cells.

The environmental tolerance of enzymes closely parallels other substances associated with live processes. They tolerate a relatively narrow temperature span and with denaturation (deactivation) occurring at temperatures generally above 50°C (122°F). Greatly reduced activity usually occurs well above the freezing point of water. Enzymes have a low tolerance to a pH below 4.0, and a minimal to no tolerance of certain organic solvents (alcohol, acetone, etc.), and destruction by numerous organic and inorganic substances.

Unlike most inorganic catalysts, enzymes are very specific for the transformations they catalyze. An acid catalyst, for example, will yield glucose, fructose, and galactose in the hydrolysis of raffinose (a trisaccharide). But, the enzyme diastase will yield melibose and fructose; emulsin will yield sucrose and galactose. The glucosidic linkages are hydrolyzed at about equal rates with an acid catalyst, whereas the enzyme catalysts act on just one kind of linkage even though the difference in linkages is small. Whereas acids may catalyze numerous compounds, including amides, acetals, and esters, a given enzyme will confine its actions to a very specific compound or closely related group. This characteristics adds very much to the efficiency of the enzymes when used in bioprocesses.

Another advantage of enzymes relates back to their microorganism precursors. Through the application of genetic recombination technology, enzyme-source microorganisms can be customized, that is, a wild bacterium or fungus can be manipulated to call for more desirable properties and the elimination of undesirable characteristics. Mutation is one way to bring this about. As pointed out by Hopwood (see reference), in *point mutation*, one can change one base pair (example: adenine-thymine to guanine-cytosine; or a base pair on a short stretch of DNA may be deleted from a sequence. Such (spontaneous) changes occur naturally, but they happen rarely (one in a million). This frequency of mutation can be multiplied hundreds of time by exposing microorganisms to mutagenic x-rays, gamma rays, or neutrons. With this technique, one can hope to find the desired mutant by examining only hundreds or a few thousands of samples instead of millions.

Suitability for Bioprocesses. In addition to the very large role that enzymes play in life processes and medicine and in industrial fermentation and related processes, enzymes are finding a growing role in industrial products, such as detergents, where enzymes tend to break down proteins to water-soluble proteoses or peptones. Enzymes for such use must remain active at relatively high pH values (8.5 to 9.5) and remain stable for a long product shelf life. See also **Detergents.**

The great number of reactions catalyzed by the enzymes in living organisms can be indicated by mentioning some of the major types. They include all the oxidation processes by which the organism obtains its energy—mechanical and thermal; the hydrolysis processes by which food carbohydrates, proteins, and fats are broken down into simpler molecules capable of direct oxidation or of use by the organism in constructing its own structure; and all the detoxification reactions by which many harmful substances that may be absorbed by the organism, as well as its normal waste products, are converted into forms suitable for excretion.

Activators. Most enzymes can function only with the assistance of certain other substances. These are broadly designated as *activators*, and are commonly grouped into two classes. The first is that of the nonspecific activators, which take no part in the conversion and appear to act by their effect upon the enzyme itself. The most important of these are the metallic ions K^+, Na^+, Rb^+, Cs^+, Mg^{2+}, Ca^{2+}, Al^{3+}, Zn^{2+}, Cd^{2+}, Cr^{2+}, Mn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+}. The second class of activators, mentioned earlier in this entry, are organic molecules, which enter into the conversion itself, often as carriers of a particular group. These substances are the group discussed in the entry on **Coenzymes.** In general, they are regenerated in their original form by other processes, so that they are not strictly substrates. On the other hand, the nicotinamideadenine nucleotides, which act as hydrogen carriers for various oxidoreductase reactions, may well be regarded as substrates.

A *substrate* may be defined as a substance modified by the action of an enzyme, or by the growing upon it of microorganisms. A *coenzyme* is a low-molecular-weight organic substance which can attach itself and thus supplement specific proteins to form active enzyme systems.

Inhibitors and Primers. Two other adjunct substances are the *inhibitors*, which retard or block enzyme action; and the *primers*, which enhance, or in some cases, are essential to it. An example is the priming of polyribonucleotide phosphotransferase by short ribonucleotide polymers.

Structure of Enzymes. The knowledge of the structure of enzymes is growing at a rapid rate, but much research remains before a high confidence level can be established pertaining to even the fundamentals of certain basic conversions.

Molecular Biology of Enzymes

One of the early examples of molecular biology which occurred after the 1950s when the DNA molecular configuration was established, but before restriction enzymes were discovered (early 1970s) was the work of researchers during the late 1950s and early 1960s on investigating the enzyme *ribonuclease.* The systematic name of this enzyme is polyribonucleoside-2-oligonucleotide-transferase. This enzyme transfers a phosphate group from one position to another within a polynucleotide, forming a cyclic compound, and the pancreatic form of this enzyme can also catalyze the transfer of the phosphate group to water, which is a step in the depolymerization of RNA.

The molecular weight of this enzyme was found (by analysis of its constituent amino acids) to be 13,700. It was found to consist of a single polypeptide chain, internally cross-linked by four cystine residues, as evidenced by the failure of any drop in molecular weight to accompany the oxidation of all four cystines to cysteic acid, and also by the occurrence of only one terminal —NH_2 group and one terminal —COOH group.

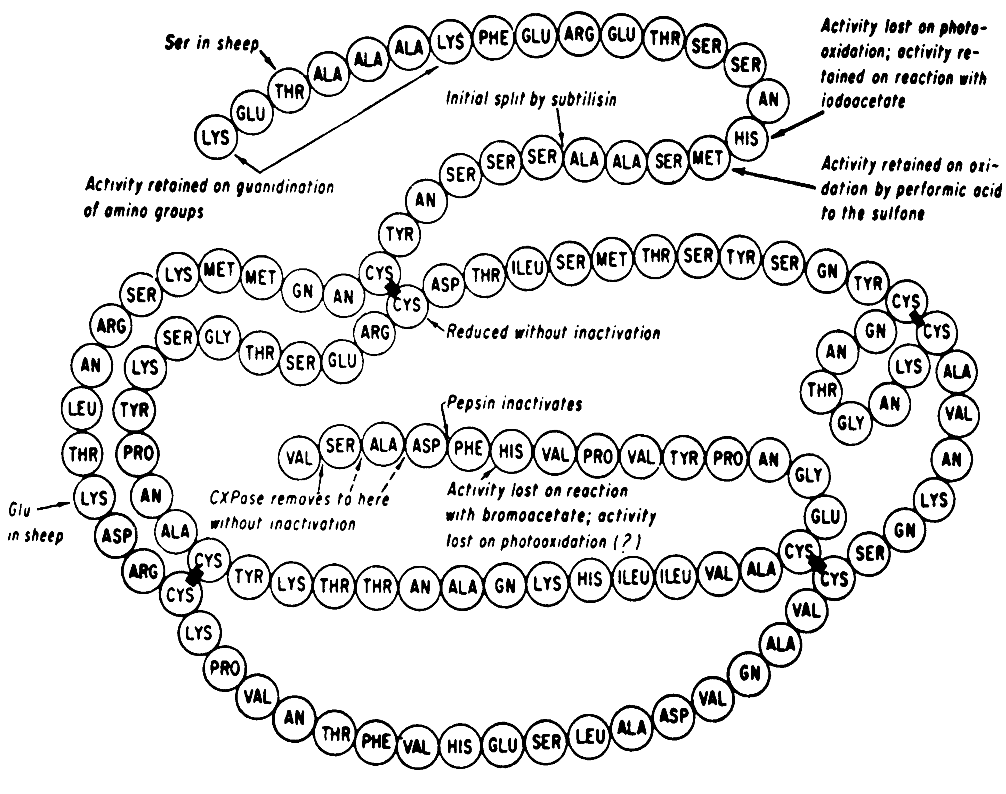
From this point, the primary structure was fully determined. The *primary structure* of an enzyme, or other protein, is the number, length, and composition of the polypeptide chains, the linear arrangement of their amino acids, and the number and position of the cross-links between chains. (The geometrical configuration of the molecule, which is usually a three-dimensional coiled and folded structure, and the side chains and their interactions were not determined.)

To determine the primary structure, after oxidation of the disulfide bridges between the cysteine molecules, the enzyme was cleaved into a series of linear polypeptides by the enzymatic action of trypsin. The fragments were separated by chromatography, and their individual amino acids were split off by acid hydrolysis. Then by repeating this process with another enzyme, a different series of polypeptide fragments were obtained, because the chains split at different points. By studying the overlaps of the two series, the fragments could be arranged in linear order. Combining the peptide structure so determined with the amino acids found gave the provisional primary structure of the enzyme shown in the accompanying figure.

The final step in complete elucidation of the three-dimensional structure of ribonuclease was made by researchers at the Roswell Park Memorial Institute, Buffalo, New York. This group, headed by Dr. David Harker, employed x-ray diffraction techniques. Roughly 500,000 diffraction points were recorded and the data so obtained was fed to a computer.

The elucidation of the structure of ribonuclease follows that of lycozyme by a group at London's Royal Institute headed by Dr. David C. Phillips, and that of the other protein, myoglobin, for which Dr. Max F. Perutz and Dr. John C. Kendrew of Cambridge University received the Nobel Prize in chemistry in 1962.

Note in this structure of enzyme that specific points are marked as those at which other enzymes act. This feature of "active sites" is char-



Early representation of ribonuclease structure. Legend: Alanine (ALA), adenine (AN), arginine (ARG), aspartic acid (ASP), cysteine (CYS), cytosine (CN), glutamic acid (GLU), glycine (GLY), guanine (GN), histidine (HIS), isoleucine (ILEU), leucine (LEU), lysine (LYS), methionine (MET), phenylalanine (PHE), proline (PRO), serine (SER), threonine (THR), thymine (TN), tryptophan (TRY), tyrosine (TYR), uracil (UN), and valine (VAL).

acteristic of enzyme behavior. In the case of the enzymes trypsin or chymotrypsin, the active sites for peptide or ester hydrolysis contain the functional groups of two histidine residues and of a serine residue. The ester enters the active region, forming temporary bonds with the enzyme at that point, the —OR group of the ester becoming bonded to a hydrogen atom of the enzyme. Then the bond between the hydrogen atom and the enzyme breaks, releasing the alcohol of the ester. As a next step, H_2O adds from the solution to the complex, and by another bond rupture, the acid part of the ester is released, leaving the enzyme in its original condition. The overall reaction is a simple hydrolysis of the ester,

$$R'COOR + H_2O \rightarrow ROH + R'COOH$$

but a large number of steps may be involved.

The determination of sites of active centers is effected not only by splitting of enzymes, but also by treating them with temporary or permanent inhibitors, and determining their points of attachment. Other methods of studying enzymes are by means of *enzyme induction* and *enzyme repression*.

Enzyme Induction. An example of *enzyme induction* is the growth of the bacteria *Escherichia coli* in a suitable culture medium. If no beta-galactoside is added to the medium, the bacteria form scarcely any of the enzyme that hydrolyzes that sugar. The addition of the sugar to the medium increases the production of the enzyme by the cell by as much as 10,000 times. On the other hand, the same bacteria will produce the enzyme tryptophan synthase only if tryptophan is absent from the culture medium. These observations are useful, not only in interpreting enzyme action, but also in determining its relationship to genetics, for in this case, the genes determining the ability to synthesize both enzymes are present on the chromosome map of the organism.

There are methods used to study enzymes other than those of chemical instrumental analysis, such as chromatography, that have already been mentioned. Many enzymes can be crystallized, and their structure investigated by x-ray or electron diffraction methods. Studies of the kinetics of enzyme-catalyzed reactions often yield useful data, much of this work being based on the Michaelis-Menten treatment. Basic to this approach is the concept that the action of enzymes depends upon the formation by the enzyme and substrate molecules of a complex, which has a definite, though transient, existence, and then decomposes into the products of the reaction. Note that this point of view was the basis of the discussion of the specificity of the active sites discussed above.

A simple enzyme reaction may thus be written as

$$E + S \leftrightarrows ES \rightarrow \text{products}$$

In the Michaelis-Menten treatment, this equation can be regarded as the result of the three processes:

rate of formation of $ES = k_1[E][S]$
rate of decomposition of ES into products $= k_2[ES]$
rate of decomposition of ES into original reactants $= k_3[ES]$

where the terms in square brackets denote concentrations of E, S, and ES, and the k's are rate constants. By representing the ratio $(k_2 + k_3)/k_1$ by K_m, the *Michaelis constant*, we can obtain a form of the Michaelis equation

$$\frac{[E_t]}{v} = \frac{K_m}{k_3 + [S]} + \frac{1}{k_3}$$

where E_t is the total concentration of enzyme (as distinguished from $[E]$, the concentration of free enzyme), and v is the velocity of the reaction. By plotting $[E_t]/v$ against $1/[S]$, we can obtain, at the axis-intercepts, values of $1/K_m$ and $1/v$.

This approach has been successfully extended to the more complex enzymatic conversions involving inhibitors, activators, and even multiple reactions in which the successive action of more than one enzyme is involved.

The observation follows that in addition to the specificity of the enzymes with respect to reaction type and to structure of the substrate, the action is also confined to a single configuration of the substrate. If the molecular structure of the substrate is unsymmetrical (asymmetric) and therefore two compounds exist—one the mirror image of the other— a specific enzyme will act upon only one of the stereoisomers. This specificity is undoubtedly due to the fact that the region of interaction on the enzyme is also asymmetric and exists in only one form or configuration. Racemic mixtures or certain substrates may sometimes be separated by making use of the fact that enzymic action will affect only one of the two forms.

Enzyme Repression. Repression as applied to biochemical reactions is a process of feedback control whereby a cell limits its production of the substances produced within it. An example that has been investigated shows the nature and mechanism by which this limitation is effected. It has been found that production of the amino acid L-isoleucine by cells of the bacterium *Escherichia coli* is repressed in the presence of an excess of the product. This excess is obtained experimentally by adding the substance to the culture medium in which the bacterium is grown. A form of the L-isoleucine is used that has been labeled with a radioactive isotope, so that the mechanism of the repression can be followed.

By this means, it has been found that the excess of L-isoleucine has two distinct effects—one that is relatively slow, and another that is rapid. The slower effect is to repress production by the cell of all the enzymes required to catalyze the series of biochemical reactions in the metabolic pathway by which the cell synthesizes L-isoleucine. The fast effect is to inhibit production of the enzyme for the first reaction in the series. This enzyme is L-threonine deaminase, which removes the amino group from L-threonine as a preliminary step to its oxidation and reamination in order to produce L-isoleucine from it.

The independent existence of these two effects was demonstrated by the discovery among mutations of *E. coli*, a mutation that exhibited only one of the effects, and of another mutation that exhibited only the other, leading to the conclusion that two distinct genes are involved in the control system.

An even more striking instance of feedback control is found in the synthesis of DNA (see **Nucleic Acids and Nucleoproteins**). As pointed out in that entry, normal DNA is composed of the nucleotides deoxyguanosine, deoxycytidine, deoxyadenosine, and thymidine, and the amounts of the first and second of these are the same, as are those of the third and fourth. Obviously, close control is required of the amounts of these nucleotides that are synthesized by the cell, if they are to be made in the quantities required for DNA synthesis. Evidence has been found that the enzyme carbamoylphosphate: L-aspartate carbamoyl transferase, which catalyzes the conversion between aspartic acid and carbamoyl phosphate (which has deoxycytidine triphosphate, CTP, as its final product), is inhibited by an excess of the CTP, and is also initiated (or activated) by an excess of deoxyadenosine triphosphate, which requires an equal amount of CTP to react with it in forming DNA. There are thus both positive and negative feedback controls on the synthesis of the enzymes that catalyze the synthesis of the nucleotides. Since all enzymes are proteins, the mechanism of this control is believed to be that suggested for protein synthesis in the entry on **Nucleic Acids and Nucleoproteins.** See also **Gene Science;** and **Molecular Biology.**

An aspect of the control of enzymatic action that is related to the effect of initiation (or activation) just discussed is the effect of *induction*, which can readily be illustrated experimentally. Many years ago, it was found that the yeast, *Saccharomyces ludwigii*, although able to ferment many sugars, was ineffective on lactose (milk sugar), because it did not synthesize the necessary enzyme, lactase (ςD-galactoside galactohydrolase). However, if this yeast was grown for several generations on a medium containing lactose, it acquired the ability to make lactase, and its subsequent generations retained that ability. In the years since this discovery, so many instances of induction have been discovered that they are regularly cited in discussions of the properties of those enzymes for which they are known, as are also the repressing and blocking substances.

Classification of Enzymes by Function

As shown by the accompanying table, enzymes are classified into six groups: (1) oxidoreductases, (2) transferases, (3) hydrolases, (4) lyases, (5) isomerases, and (6) ligases or synthetases. The main group to which an enzyme belongs is indicated by the first figure of the code number. The second figure indicates the subclass; for the oxidoreductases, it shows the type of group in the *donors* which undergoes oxidation; for

CLASSIFICATION OF ENZYMES

1. Oxidoreductases
   - 1.1 *Acting on the* CH—OH *group of donors*
     - 1.1.1 With NAD or NADP as acceptor
     - 1.1.2 With cytochrome as an acceptor
     - 1.1.3 With O_2 as acceptor
     - 1.1.99 With other acceptors
   - 1.2 *Acting on the aldehyde or keto group of donors*
     - 1.2.1 With NAD or NADP as acceptor
     - 1.2.2 With a cytochrome as an acceptor
     - 1.2.3 With O_2 as acceptor
     - 1.2.4 With lipoate as acceptor
     - 1.2.99 With other acceptors
   - 1.3 *Acting on the* CH—CH *group of donors*
     - 1.3.1 With NAD or NADP as acceptor
     - 1.3.2 With a cytochrome as an acceptor
     - 1.3.3 With O_2 as acceptor
     - 1.3.99 With other acceptors
   - 1.4 *Acting on the* CH—NH_2 *groups of donors*
     - 1.4.1 With NAD or NADP as acceptor
     - 1.4.3 With O_2 as acceptor
   - 1.5 *Acting on the* C—NH *group of donors*
     - 1.5.1 With NAD or NADP as acceptor
     - 1.5.3 With O_2 as acceptor
   - 1.6 *Acting on reduced* NAD *or* NADP *as donor*
     - 1.6.1 With NAD or NADP as acceptor
     - 1.6.2 With a cytochrome as an acceptor
     - 1.6.4 With a disulfide compound as acceptor
     - 1.6.5 With a quinone or related compound as acceptor
     - 1.6.6 With a nitrogenous group as acceptor
     - 1.6.99 With other acceptors
   - 1.7 *Acting on other nitrogens compounds as donors*
     - 1.7.3 With O_2 as acceptors
     - 1.7.99 With other acceptors
   - 1.8 *Acting on sulfur groups of donors*
     - 1.8.1 With NAD or NADP as acceptor
     - 1.8.3 With O_2 as acceptor
     - 1.8.4 With a disulfide compound as acceptor
     - 1.8.5 With a quinone or related compound as acceptor
     - 1.8.6 With a nitrogenous group as acceptor
   - 1.9 *Acting on heme groups of donors*
     - 1.9.3 With O_2 as acceptor
     - 1.9.6 With a nitrogenous group as acceptor
   - 1.10 *Acting on diphenols and related substances as donors*
     - 1.10.3 With O_2 as acceptor
   - 1.11 *Acting on* H_2O_2 *as acceptor*
   - 1.12 *Acting on hydrogen as donor*
   - 1.13 *Acting on single donors with incorporation of oxygen (oxygenases)*
   - 1.14 *Acting on paired donors with incorporation of oxygen into one donor (hydroxylases)*
     - 1.14.1 Using reduced NAD or NADP as one donor
     - 1.14.2 Using ascorbate as one donor
     - 1.14.3 Using reduced pteridine as one donor
2. Transferases
   - 2.1 *Transferring one-carbon groups*
     - 2.1.1 Methyltransferases
     - 2.1.2 Hydroxymethyl-, formyl-, and related transferases
     - 2.1.3 Carboxyl- and carbamoyltransferases
     - 2.1.4 Amidinotransferases
   - 2.2 *Transferring aldehydic or ketonic residues*
   - 2.3 *Acyltransferases*
     - 2.3.1 Acyltransferases
     - 2.3.2 Aminoacyltransferases
   - 2.4 *Glycosyltransferases*
     - 2.4.1 Hexosyltransferases
     - 2.4.2 Pentosyltransferases
   - 2.5 *Transferring alkyl or related groups*
   - 2.6 *Transferring nitrogenous groups*
     - 2.6.1 Aminotransferases
     - 2.6.3 Oximinotransferases
   - 2.7 *Transferring phosphorus-containing groups*
     - 2.7.1 Phosphotransferases with an alcohol group as acceptor
     - 2.7.2 Phosphotransferases with a carboxyl group as acceptor
     - 2.7.3 Phosphotransferases with a nitrogenous group as acceptor
     - 2.7.4 Phosphotransferases with a phospho-group as acceptor
     - 2.7.5 Phosphotransferases, apparently intramolecular
     - 2.7.6 Pyrophosphotransferases
     - 2.7.7 Nucleotidyltransferases
     - 2.7.8 Transferases for other substituted phospho-groups
   - 2.8 *Transferring sulfur-containing groups*
     - 2.8.1 Sulfurtransferases
     - 2.8.2 Sulfotransferases
     - 2.8.3 CoA-transferases
3. Hydrolases
   - 3.1 *Acting on ester bonds*
     - 3.1.1 Carboxylic ester hydrolases
     - 3.1.2 Thiolester hydrolases
     - 3.1.3 Phosphoric monoester hydrolases
     - 3.1.4 Phosphoric diester hydrolases
     - 3.1.5 Triphosphoric monoester hydrolases
     - 3.1.6 Sulfuric ester hydrolases
   - 3.2 *Acting on glycosyl compounds*
     - 3.2.1 Glycoside hydrolases
     - 3.2.2 Hydrolyzing N-glycosyl compounds
     - 3.2.3 Hydrolyzing S-glycosyl compounds
   - 3.3 *Acting on ether bonds*
     - 3.3.1 Thioether hydrolases
   - 3.4 *Acting on peptide bonds (peptide hydrolases)*
     - 3.4.1 α-Aminoacyl-peptide hydrolases
     - 3.4.2 Peptidyl-amino acid hydrolases
     - 3.4.3 Dipeptide hydrolases
     - 3.4.4 Peptidyl-peptide hydrolases
   - 3.5 *Acting on* C—N *bonds other than peptide bonds*
     - 3.5.1 In linear amides
     - 3.5.2 In cyclic amides
     - 3.5.3 In linear amidines
     - 3.5.4 In cyclic amidines
     - 3.5.5 In cyanides
     - 3.5.99 In other compounds
   - 3.6 *Acting on acid-anhydride bonds*
     - 3.6.1 In phosphoryl-containing anhdrides
   - 3.7 *Acting on* C—C *bonds*
     - 3.7.1 In ketonic substances
   - 3.8 *Acting on halide bonds*
     - 3.8.1 In C-halide compounds
     - 3.8.2 In P-halide compounds
   - 3.9 *Acting on* P—N *bonds*
4. Lyases
   - 4.1 *Carbon-carbon lyases*
     - 4.1.1 Carboxyl-lyases
     - 4.1.2 Aldehyde-lyases
     - 4.1.3 Ketoacid-lyases
   - 4.2 *Carbon-oxygen lyases*
     - 4.2.1 Hydro-lyases
     - 4.2.99 Other carbon-oxygen lyases
   - 4.3 *Carbon-nitrogen lyases*
     - 4.3.1 Ammonia-lyases
     - 4.3.2 Amidine-lyases
   - 4.4 *Carbon-sulfur lyases*
   - 4.5 *Carbon-halide lyases*
   - 4.99 *Other lyases*
5. Isomerases
   - 5.1 *Racemases and epimerases*
     - 5.1.1 Acting on amino acids and derivatives
     - 5.1.2 Acting on hydroxyacids and derivatives
     - 5.1.3 Acting on carbohydrates and derivatives
     - 5.1.99 Acting on other compounds
   - 5.2 *Cis-trans isomerases*
   - 5.3 *Intramolecular oxidoreductases*
     - 5.3.1 Interconverting aldoses and ketoses
     - 5.3.2 Interconverting keto- and enol-groups
     - 5.3.3 Transposing C??C bonds
   - 5.4 *Intramolecular transferases*
     - 5.4.1 Transferring acyl groups
     - 5.4.2 Transferring phosphoryl groups
     - 5.4.99 Transferring other groups
   - 5.5 *Intramolecular lyases*
   - 5.99 *Other isomerases*
6. Ligases or Synthetases
   - 6.1 *Forming* C—O *bonds*
     - 6.1.1 Aminoacid-RNA ligases
   - 6.2 *Forming* C—S *bonds*
     - 6.2.1 Acid-thiol ligases
   - 6.3 *Forming* C—N *bonds*
     - 6.3.1 Acid-ammonia ligases (amide synthetases)
     - 6.3.2 Acid-amino acid ligases (peptide synthetases)
     - 6.3.3 Cyclo-ligases
     - 6.3.4 Other C—N ligases
     - 6.3.5 C—N ligases with glutamine as N-donor
   - 6.4 *Forming* C—C *bonds*

the transferases, it indicates the nature of the group which is transferred; for the hydrolases, it shows the type of bond hydrolyzed; for the lyases, the type of link which is broken between the group removed and the remainder; for the isomerases, the type of isomerization involved; and for ligases, the type of bond formed.

The third figure of the code number, indicating the sub-sub class, shows for the oxidoreductases the type of acceptor involved; for the transferases and hydrolases, it shows more precisely the type of group transferred or bond hydrolyzed; for the lyases, it shows the nature of the group removed; for the isomerases, it indicates in more detail the nature of the isomerization; and for the ligases, it shows the nature of the substance formed. Thus, an enzyme number, commonly indicated by the prefix EC, provides fairly detailed information about a specific enzyme.

Categories of Conversions. The comparative simplicity of the classification scheme bears testimony to the underlying unity of enzymatic catalysis.

Oxidoreductases. The overall conversion catalyzed by the oxidoreductases can be written as hydrogen transfer, and these enzymes might be considered to be merely one section of the transferases. The oxidoreductases are classified separately because of their large number and because of their great biological importance in bringing about the main energy-yielding conversions of living tissues.

Transferases. The main groups of transferases are concerned with the transfer of *one-carbon* groups, acyl groups, glycosyl residues, amino- and other nitrogen-containing groups, phosphate, and sulfate. Oxidoreductases and transferases together represent about half or more of the enzymes presently recognized. A general conversion for both oxidoreductases and transferases can be written:

$$AX + B \rightleftharpoons A + BX$$

Hydrolases. These enzymes include esterases, glycosidases, peptidases, deaminases, and enzymes which hydrolyze acid anhydrides (such as the pyrophosphate group in adenosinetriphosphate). Many hydrolases have been shown to be able, under appropriate conditions, to catalyze transfer conversions; a high concentration of acceptor is usually necessary, since there is competition between the added acceptor and water for the group transferred. The detailed mechanism in these cases probably involves transfer of a part of the substrate onto a group on the enzyme, with subsequent transfer to an acceptor or hydrolysis, e.g., for a hydrolase acting on a substrate AB to produce AOH and BH:

$$EH + AB \rightarrow E - A + BH$$

and

$$E - A + X \rightarrow E + AX$$

or

$$E - A + H_2O \rightarrow EH + AOH$$

These hyrolases, if not all, can therefore be regarded as transferases which include H_2O among their possible acceptors. Under normal conditions, in aqueous solution, hydrolysis will be the dominant conversion.

Lyases. Enzymes in this grouping catalyze conversions of the type:

$$AX - BY \rightleftharpoons A = B + X - Y$$

Molecules, such as H_2O, H_2S, NH_3, or aldehydes, are added across the double bond of a second unsaturated molecule. Decarboxylases, such as those acting on amino acids can be regarded as lyases (carboxylyases), assuming CO_2 and not H_2CO_3 to be the immediate product of decarboxylation. Over one hundred lyases are known.

Isomerases. These include enzymes which bring about conversions similar to those in several other groups, but distinguished in that the reaction takes place entirely within one molecule, which is not cleaved, so that the overall reaction is

$$A \rightleftharpoons B$$

Thus, there are intramolecular oxidoreductases (e.g., ketolisomerases), intramolecular transferases (e.g., phosphomutases), and intramolecular lyases. About fifty isomerases are known.

Ligases. These enzymes catalyze conversions which are more complex than those of the other groups and must involve at least two separate stages in the reaction. The overall result is the synthesis of a molecule from two components with a coupled breakdown of adenosine

$$X + Y + ATP \rightarrow XY + AMP + \text{Pyrophosphate (or ADP + Phosphate)}$$

triphosphate, or some other nucleoside triphosphate. In general, this may be written:

These enzymes, of which many are known, are of great importance in the conservation of chemical energy within the cell and in the coupling of synthetic processes with energy-yielding breakdown conversions.

Additional Reading

Beardsley, T. M.: "Molecular Engineers Mimic Mother Nature: Synthetic Enzymes," *Sci, Amer.*, 30 (September 1990).

Dziezak, J. D.: "Enzymes: Catalysts for Food Processes," *Food Technology*, 77 (January 1990).

Fisckher, E. H., Charbonneau, H., and N. K. Tonkas: "Protein Tyrosine Phosphatases: A Diverse Family of Intracellular and Transmembrane Enzymes," *Science*, 401 (July 26, 1991).

Gross, A.: "Enzymatic Catalysis in the Production of Novel Food Ingredients," *Food Technology*, 96 (January 1991).

Holsinger, V. H., and A. E. Kligerman: "Applications of Lactase in Dairy Foods and Other Foods Containing Lactose," *Food Technology*, 92 (January 1991).

Kennedy, J. F., and E. H. M. Melo: "Immobilized Enzymes and Cells," *Chem. Eng. Progress*, 81 (July 1990).

Kornberg, A.: "For the Love of Enzymes: The Odyssey of a Biochemist," Harvard University Press, Cambridge, Massachusetts, 1989.

Liebman, J. F., and A. Greenberg: "Mechanistic Principles of Enzyme Activity," VCH Publishers, New York, 1989.

Neidlemann, S. L.: "Enzymes in the Food Industry," *Food Technology*, 88 (January 1991).

Nielsen, H. K.: "Novel Bacteriolytic Enzymes and Cyclodextrin Glycosyl Transferase for the Food Industry," *Food Technology*, 102 (January 1991).

Penet, C. S.: "Applications of Industrial Food Enzymology," *Food Technology*, 98 (January 1991).

Staff: "Industrial Food Enzymology," *Food Technology*, 87 (January 1991).

Stauffer, C. E.: "Enzyme Assays for Food Scientists," Van Nostrand Reinhold, New York, 1989.

Wasserman, B. P.: "Evolution of Enzyme Technology: Progress and Prospects," *Food Technology*, 118 (April 1990).

ENZYME PREPARATIONS. During the past several years, a number of commercially prepared enzyme preparations have been available to processors, notably for use in the food industry. These preparations fall into three basic categories: (1) Animal-derived preparations; (2) plant-derived preparations; and (3) microbially derived preparations.

Fruit juices, jams, and jellies, corn (maize) syrups and sweeteners, structured protein foods, and tenderized meats are exemplary of products, the quality of which has been improved through the use of enzyme preparations. Principal areas of development in food-grade enzyme research have been toward upgrading quality and byproduct utilization, higher rates and levels of extractions, synthetic food development, sweetener development, improving flavor of foods, and the stabilization of food quality and nutrition. Enzyme preparations also are used in the detergent field.

Animal-derived enzyme preparations include catalase (bovine liver), lipase, pepsin, rennet, and trypsin. Plant-derived preparations include bromelain, cellulase, ficin, malt, papain, and pectinase. Microbially derived preparations include amylases, carbohydrase, catalase, glucose oxidase, lipase, protease, and zymase.

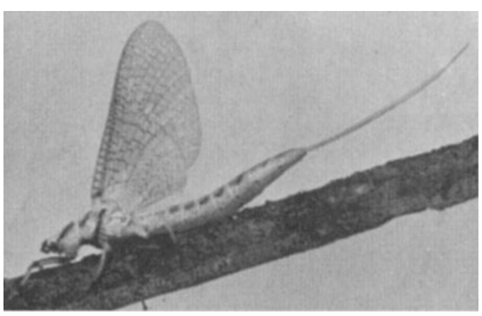
EOCENE. A subdivision of the Tertiary of the geologic time-scale. Type locality, near Paris, France. Term first proposed by Lyell in 1832. The Eocene began approximately 60 million years ago and lasted for about 20–30 million years. The greatest thickness of formations of this period occur in Wyoming. The principal areas of deposition in the United States are: (1) the unconsolidated marine gravels, glauconite sands and clays which overlap the Cretaceous marine sediments of the Atlantic Border; (2) the marine limestones, terrestrial sandstones and lignites of the Gulf Coast; (3) the marine sediments of the Pacific Coast; (4) the terrestrial intermontane deposits of the Western interior.

The plants of this period suggest worldwide warm climate. The fossil plants include many of the modern genera, such as the beeches, dogwoods, walnuts, maples and elms. Fossil vertebrate skeletons show that the mammals were by then dominant, although many of the existing orders of reptiles and birds also lived at this time. The mammalian fauna may be divided into two principal groups: (1) the archaic types which did not survive the Eocene; (2) the progenitors of the modern mammals, including the ancestors of the camels, pigs, horses, rats, and primitive monkeys. The principal surviving archaic forms are the creodonts (primitive flesh-eaters), uintatheria (hippopotamus-like forms), and zeuglodons (marine mammals). The mineral resources of this period are described under **Tertiary**.

EOLIAN DEPOSITS. Sediments and sedimentary rocks which are largely, if not entirely, composed of wind-blown material. Desert sands are typical aeolian sediments, characterized by relatively uniform, well-rounded particles whose surfaces are usually covered with microscopic pits due to their mutual bombardment during transportation. This pitting gives each sand grain a frosted appearance. Wind-blown sediments frequently show characteristic cross-bedding, ripple marks (miniature dunes) and wind-faceted pebbles (glyptoliths). Further evidence of their origin is the absence of fossils. Aeolian deposits are usually largely composed of quartz sand. An important fine-grained wind-blown deposit is loess. Extensive desert deposits are also composed of gypsum, salt, etc.

EOPHYTIC. A paleobotanic division of geologic time. This term signifies the time during which algae were abundant. See also **Paleobotany.**

EPHEMEROPTERA. The mayflies, also known locally as shad flies, salmonflies and June bugs. The adults are sluggish insects with slender filaments at the caudal end of the body and large triangular front wings. The hind legs are much smaller, in some species rudimentary. The immature insect is aquatic and in most species feeds on decaying vegetable matter. It may live for several years, while the adult stage lasts only a few days.

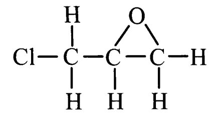


Mayfly (*Ephemeroptera*). (*A. M. Winchester.*)

Mayflies of one species emerge as adults in large numbers within a short period and are sometimes very abundant near favorable bodies of water. They fly at twilight and can sometimes be seen in large gray clouds at a distance of more than a mile over the islands of Lake Erie, where they are especially abundant. In the cities bordering the lake they are attracted to lights and their dead bodies are sometimes swept up in bushels after a heavy flight. Under such conditions they are a nuisance but not a serious pest. Their value as food for fishes more than offsets what little harm they do.

EPICARDIUM. 1. The thin covering of the vertebrate heart, continuous with the lining of the pericardial cavity. 2. Outgrowth from the branchial sac in many ascidians, which takes part in budding.

EPICHLOROHYDRIN. A highly reactive and industrial important compound with the structural formula



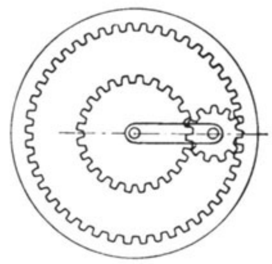
It is also called 1-chloro-2,3-epoxypropane and is classified as an organic epoxide. The compound is a colorless, clear, mobile liquid with an odor something like chloroform. Molecular weight, 92.53; freezing point, −57.1°C; boiling point, 116.07°C; density 1.1750 g/cm^3 at 25°C with reference to water at 4°C. Solubility is 6.53 g/100 g of water. Epichlorohydrin is made by the chlorohydrination of allyl chloride, in which 1,2-dichlorohydrin and 1,3-dichlorohydrin are produced as intermediates.

One of the most common epoxy resins is produced by the reaction between epichlorohydrin and bisphenol A. See also **Epoxy Resins.** The compound also is used in the production of epichlorohydrin-based rubbers which have good aging, high resiliency, and flexibility at low temperatures, advantage of which is taken in automotive and aircraft parts, seals, gaskets, hose, belting, wire, and cable jackets. These rubbers also have good resistance to solvents, fuels, oils, and ozone. A number of wet-strength resins for use in the paper industry also are derived from epichlorohydrin, including (a) epichlorohydrin-modified polyamides; and (b) the addition of epichlorohydrin to high-molecular-weight polyalkylene polyamines. The advantages of these resins is that no alum or acid medium is required for incorporating the resin into the cellulose pulp. During the drying process, the resin cross-links and thus yields a paper with permanent wet-strength properties. Ion-exchange resins also can be prepared by reacting epichlorohydrin with ethylene diamine or a similar amine. The resulting material is a stable, water-insoluble anion-exchange resin.

In addition to its use in the production of epoxy resins epichlorohydrin is used in large quantities in the manufacture of glycerin. Other uses include textile applications where it is used to modify the carboxy groups of wool, thus increasing durability and improving moth resistance; in the synthesis of antistatic agents, wrinkle-resistant agents, and coating sizings. Effective against the larvae of certain insects, the compound is used in control chemicals for agriculture where permitted.

EPICONTINENTAL SEAS (or Epicontinental Marginal Seas). Shallow bodies of water deeper than continental shelves and having somewhat greater relief. Generally somewhat greater than 600 feet (180 meters) in depth. See **Continental Shelves.**

EPICYCLIC GEAR TRAIN. Combinations of gears having a motion resulting from rotation about an axis which, in itself, is in rotation, are known as epicyclic trains. A simple epicyclic gear train, consisting of three gears and an arm, is shown in the figure. Mechanism of this nature is sometimes used for speed reducers. The ratios of speed of the driven and driving elements are found by the following simple rule: consider, first, the gears locked and the entire mechanism turned one revolution; then the arm locked and the fixed gear turned one revolution in the opposite direction to the first step. The algebraic sum of these two separate motions will give the absolute number of turns of any gear, and from this the speed ratio may be found.



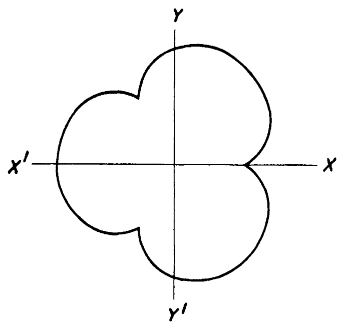
Epicyclic gear train.

EPICYCLOID. A higher plane curve, which is a special case of a cyclic curve. A circle of radius r rolls around the outside of a fixed circle of radius R and a point on the circumference of the moving circle traces out the curve. Its equation in parametric form is

$$x = (R + r)\cos\phi - r\cos\frac{(R + r)\phi}{r}$$

$$y = (R + r)\sin\phi - r\sin\frac{(R + r)\phi}{r}$$

The curve consists of a set of congruent arches. The first one will occur at some value of ϕ between 0 and $2\pi r/R$ and the kth one between $2\pi r(k - 1)/R$ and $2\pi kr/R$. It will not repeat itself, however, unless $2\pi kr/R$ is a multiple of 2π for k, an integer. This means that $r/R = m/n$, where m/n is a rational fraction. Thus $k = n$, there are n arches to the curve, an equal number of cusps, and it winds around the fixed circle m times.



Epicycloid.

The special case of $r = R$ is called a cardioid. If the generating point for the epicycloid is at a distance $a \neq r$, the curve is an epitrochoid.

If an epicycloid rolls on a straight line, the locus of the center of the fixed circle is an ellipse, which is called the roulette of the epicycloid. The evolute of an epicycloid is a similar epicycloid.

See also **Cardioid.**

EPIDEMIOLOGY. That branch of medical science which is concerned with the study of disease as it appears in its natural surroundings, and as it affects a community of people rather than a single individual. Epidemiology is largely concerned with infectious diseases; it has a statistical as well as an experimental side. The statistical includes the gathering of facts about the incidence, mortality rate, relation of climate, age, sex, race and many other factors to the appearance of a given disease. From these data, valuable information which is important in controlling epidemics of disease is obtained. In experimental epidemiology, the investigator may produce epidemics in laboratory animals for the study of certain problems, or he may go out in the field and study epidemics in man by laboratory means. Such work is important in elucidating the cause, the modes of transmission, and the insect vectors possibly concerned, as well as other facts of fundamental importance in public health. Numerous governments throughout the world maintain statistical and data processing centers which provide periodic reports on many aspects of diseases. Most states and provinces maintain their own records and also report and exchange information with a central government office, such as the Centers for Disease Control of the U.S. Public Health Service, located in Atlanta, Georgia.

Epidemiology in connection with foodborne diseases is discussed in some detail in the entry on **Foodborne Diseases.**

R. C. V.

EPIDERMIS. In insects, the outer layer of the noncellular cuticula; here the cellular layer is called the hypodermis. In other invertebrates the cellular layer covering the body is called the epidermis. In vertebrates, the epidermis is the outer cellular layer of skin. The human epidermis consists of four layers, from without to within follows: (1) a layer of horny flattened cells, (2) a layer of transparent cells, (3) layers of granular cells, (4) a layer of rounded pigmented cells. The outermost layer of cells on the younger parts of plants is also called the epidermis. See also **Dermatitis and Dermatosis.**

EPIDIORITE. A term applied to gabbros, dolerites, and diabases, the augite of which has been partly altered to hornblende, thus approaching a diorite in mineral composition. The term is derived from the Greek, meaning upon, plus diorite.

EPIDOTE. This mineral is a hydrous silicate of calcium, aluminum, and iron with the formula, $Ca_2(Al, Fe)_3Si_3O_{12}(OH)$. The ratio of aluminum to iron ranges from 6:1 to 3:2. Epidote is found in prismatic monoclinic crystals, which may be acicular to fibrous. Fine granular and compact masses are common. The mineral displays one good cleavage, an uneven fracture; is brittle; hardness, 6–7; specific gravity, 3.25–3.5; luster, vitreous to resinous; typical color, pistachio green, but may be yellowish- to brownish-green, sometimes red, yellow, gray, white or colorless. Colorless to grayish streak; transparent to opaque. The characteristic color of ordinary epidote makes it usually an easily identified mineral.

It occurs commonly in metamorphic rocks as gneisses and schists; however, it seems probable that under certain conditions it may appear as a primary mineral, for example in granitic rocks. The Urals, Austria, Switzerland, Italy, France and Norway are known for their occurrences of fine epidote crystals. In the United States epidote has been found in excellent specimens at Franconia and Warren, New Hampshire; Huntington, Massachusetts; Willimantic and Haddam, Connecticut; Chaffee County, Colorado, and Riverside County, California. The word epidote is derived from the Greek. The name pistacite, from the Greek word meaning pistachio nut, has been occasionally applied to this mineral. It has been used as a gemstone but is in little demand for this purpose.

EPIGENETIC. A term used by petrologists to denote physical and chemical changes, particularly in igneous and sedimentary rocks, which are clearly secondary to (later in time) the conditions under which the rock originated. This term is commonly used by the students of ore deposits to designate minerals formed after the enclosing wall rocks, in contrast to those minerals formed contemporaneously with the wall rocks. The latter minerals are said to be syngenetic. In the case of the sedimentary rocks, the term is used to describe textures, structures and mineral aggregates, of nonmetamorphic origin, which have originated during the postlithification history of the formation. Thus flint, chert, and concretions, may be described as being either epigenetic or syngenetic.

EPIGLOTTITIS. The pharynx is a common passageway of the respiratory and digestive systems. See **Pharynx.** A valvelike structure at the base of the tongue, the *epiglottis*, projects backward over the larynx during swallowing, and thereby prevents food from entering the larynx. Acute *epiglottitis*, an infection of this valve, is one of the most rapidly progressive and sometimes more lethal than any of the other infections of the upper respiratory tract—particularly among children between ages of 2 and 8 years. The condition starts suddenly, with very severe sore throat and fever, leading to dysphagia (difficulty in swallowing). There may be retention of secretions and drooling. Respiratory obstruction may occur within a few hours. The condition may occur in adults, but progresses more slowly because of the larger size of the airway in an adult.

Acute epiglottitis is considered a medical emergency. Inasmuch as examination of the pharynx may not produce an accurate diagnosis and because tongue depressors may elicit spasms, an immediate lateral-view x-ray of the neck should be made. Frequently, swelling of the epiglottis will be indicated. Where the x-ray does not satisfactorily confirm the condition, indirect laryngoscopy may be undertaken.

Tracheostomy is the usual means taken to provide a restoration of the airway. Nasotracheal intubation is also used. To avoid delays, direct laryngoscopy may be undertaken. An important element of diagnosis is that of distinguishing epiglottitis from croup caused by viral laryngitis or other laryngeal conditions, and other possible causes of airway obstruction. In the case of croup, the epiglottis will be normal or only mildly inflamed.

The principal cause of acute epiglottitis is *H. influenzae* Type B. However, other pathogens may be involved. They include pneumococci, streptococci, and staphylococci. Physicians stress the need for immediate action in these cases, the initial therapy usually consisting of high-dose intravenous chloramphenicol (but considering its side effects) and penicillinase-resistant penicillin. Steroids may be administered to reduce edema and a mist tent may be a helpful supportive measure to assist breathing. In about half of the cases of this disease, nasotracheal intubation or tracheostomy is required.

EPIPELAGIC ZONE. The uppermost and very shallow layer of water in the oceans into which sufficient light passes to enable phytoplankton to convert the available carbon dioxide into food by means of photosynthesis. This zone, rarely more than 700 feet (210 meters) in depth, is the habitat of the major part of the sea's living matter.

EPIPHYSIS. An area of cartilage near the ends of long bones which ossifies separately and later becomes part of the long bone; bone growth in length takes place in this area until ossification halts growth. Also called *epiphyseal cartilage*. See also **Bone.**

EPIPHYTES. A striking feature of tropical forests is the abundance of plants which grow attached to other plants. These attached plants are called epiphytes, which means plants growing on other plants. They are found both on the main trunk and on the branches, often far above the ground. In many cases the epiphyte grows on the under side of a branch to which it is firmly fastened by its roots. Epiphytes gain nothing but support, and a more favorable position of growth because of better light conditions and other environmental factors; they do not obtain any nutrients from the supporting plants, as parasites would. Particularly noteworthy among epiphytes are many ferns, aroids and especially orchids. Frequently, as in the orchids, the roots of the plants are so modified as to absorb water directly from the atmosphere.

In the strictest meaning of the word, many lower plants including algae, fungi, lichens and mosses are epiphytes, since they are often found growing on other plants, not only in tropical regions but in temperate regions as well.

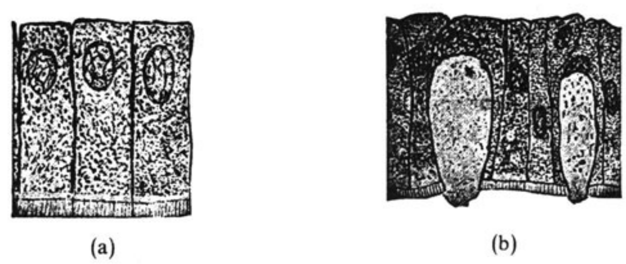
EPISTASIS. The inhibition by a gene at one locus on a chromosome of the expression of a gene at another locus. This form of suppression should be distinguished from simple dominance, which is associated with members of allelic pairs. The gene which does the inhibiting is said to be epistatic to the gene which is inhibited. The gene which is suppressed is said to be hypostatic. For instance, the gene for albinism is epistatic to the gene for black coat color in guinea pigs. A guinea pig may be homozygous for the dominant gene for black coat, but if he is also homozygous for the recessive gene for albinism, the coat is white. This is known as recessive epistasis, since two genes for albinism must be present. In dominant epistasis, only one gene is necessary to cause the inhibition. In man there is a dominant gene for dwarfism (Chondrodystropic dwarfism) and a single gene for this condition can be epistatic to the genes for normal growth.

EPISTAXIS. Hemorrhage from the nose. Nosebleed.

EPITAXY. Oriented intergrowth between two solid phases. The surface of one crystal provides, through its lattice structure, preferred positions for the deposition of the second crystal.

EPITHELIUM. Tissue which covers surfaces and lines hollow organs, and the derivatives of these tissues, whether solid or hollow. All epithelial tissues are made up of closely associated cells with very little intercellular material and most of them have one surface free and the other connected with an underlying tissue.

Epithelia are classified according to the form of cells, the number of layers, and the embryonic origin and location. See accompanying illustration. In flat, pavement, or squamous epithelium the cells are much thinner than their diameter. Cuboidal epithelium is made up of cells approximately as thick as their width. They are not strictly cuboidal but are polyhedral prisms. Columnar epithelium contains cells that are much higher than their diameter. Glandular epithelia are usually of the two latter forms. Any of these forms of cells may occur in more than one layer as a stratified epithelium, but flat epithelia are more often stratified and thicker cells usually form a single layer, or a simple epithelium. In some cases, the epithelium is made up of cells of several forms in two or three layers, which change movements of the part. Such a tissue is called traditional. Others appear to have several layers of cells but all are attached to the underlying tissue, rising to various heights. This is a pseudo-stratified epithelium.



Forms of epithelium: (a) Simple columnar epithelium from intestinal lining; (b) goblet cells from epithelium lining large intestine. (*Kimber and Gray, "Textbook of Anatomy and Physiology," Macmillan.*)

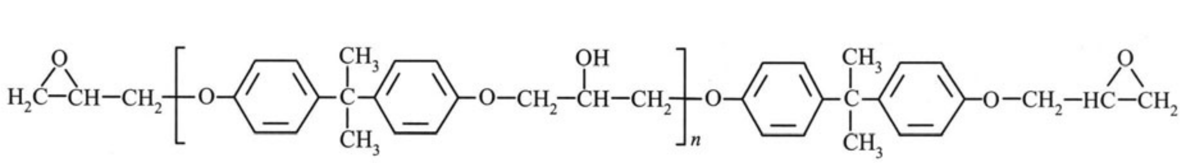
According to origin and position, two kinds of epithelia derived from the mesoderm are recognized. Of these, endothelium lines the circulatory organs and mesothelium lines the body cavity. Other special types of epithelium are derived from each of the three germ layers. The free surfaces of some bear cilia.

The cells of many epithelia produce special secretions, and in some cases, these glandular layers are highly developed to form massive structures known as glands.

Most epithelia rest on a thin basement membrane or *membrana propria* derived from the connective tissues.

EPIZOOTIC. Term descriptive of any disease in animals whose incidence and distribution resemble those of an epidemic in humans.

EPOXY RESINS. A family of thermosetting resins known for their excellent mechanical and electrical properties, dimensional stability, resistance to high temperatures and numerous chemicals, and for their strong adhesion to glass, metal, fibers, and numerous other materials. Structurally, the epoxy groups are three-membered rings with one oxygen and two carbon atoms. The most common epoxy resins are made by reacting epichlorohydrin with a polyhydroxy compound, such as bisphenol A, in the presence of a catalyst. Epoxy resins produced in this fashion are known as diglycidyl ethers of bisphenol A (bis-A). The structural formula is:



By changing the ratio of epichlorohydrin to bis-A, resins range from low-viscosity liquids to high-melting solids. The structure shown represents a solid epoxy novolak resin. The epoxy phenol novolak resins are the most important. Basically they are novolak resins whose phenolic hydroxyl groups have been converted to glycidyl ethers. Epoxidized novolaks are used principally in solid single-stage molding compounds and high-temperature laminating systems.

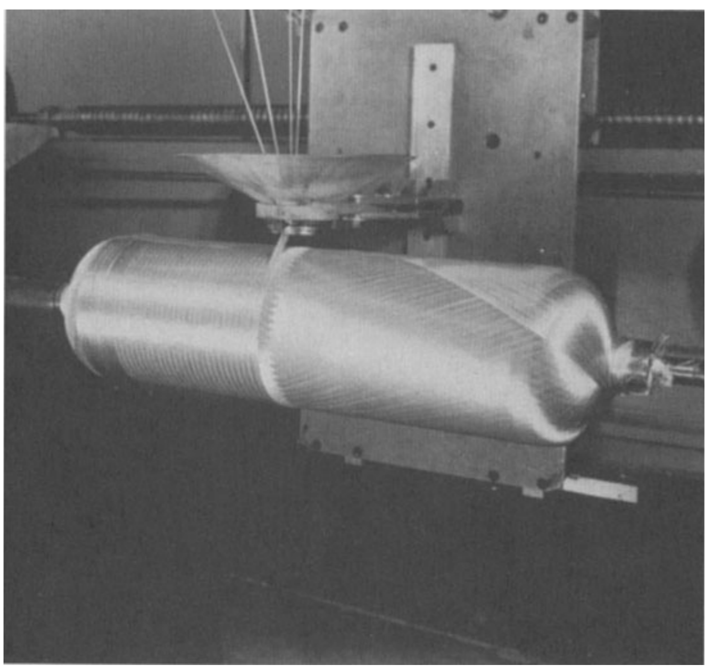
The liquid cycloaliphatic epoxies normally are produced by the peracetic acid epoxidation of cyclic olefins where the epoxide groups are attached directly to the cycloaliphatic ring. These materials have better weatherability, arc and tracking resistance and dielectric strength over conventional epoxies.

Usually the major makers of resins, hardeners and other chemicals for epoxy systems do not supply finished compounds. The compounding is done by specialized firms and by some large epoxy users. The epoxy resins per se are not finished products, but are reactive chemicals to be combined with other chemicals to yield systems capable of conversion to a predetermined thermoset structure.

Epoxy resins are cured by cross-linking agents known as hardeners or by catalysts which promote self-polymerization. Some of the cross-linking agents used include the primary and secondary aliphatic polyamines, such as diethylenetriamine, triethylenetetramine, tetraethylenepentamine, diethylaminopropylamine, and piperazines. Most of these materials are liquids of moderately low viscosity that can be blended with the resins at room temperature.

Because of excellent electrical properties, epoxies are used in casting, potting, and encapsulation of electrical/electronic parts. Advantage is taken of the low shrinkage of the epoxies, with absence of cracking or separation of the resins from the parts during cure. Encapsulated parts vary from miniature coils and switches that may weigh but a few grams to large motors and insulators that may weigh several pounds.

Epoxy resins also find use in making chemical-loaded molecular sieves, adhesives, and protective coatings. Epoxy-based adhesives are widely used for bonding dissimilar materials like plastics and metal, wood and metal, and ceramics and rubber. Minimum pressure is required to obtain a satisfactory bond. For such applications, epoxy adhesives are available as one- or two-part systems. The one-part systems require curing at elevated temperatures. The two-part systems can be cured at room temperature, but both have better resultant properties if cured under heat. Epoxy resin-based coatings provide outstanding chemical resistance, toughness, flexibility, and adhesion to most substrates.



Glass and other fibers coated with epoxy resin are filament-wound into containers of high strength. (*Union Carbide Corp.*)

Epoxy resins are used in the chemical industry because of their excellent resistance to attack by many corrosive chemicals. One application is a protective coating for container, pipe, and tank liners, floors, and walls. High-pressure vessels are made by filament winding with cycloaliphatic epoxies. See accompanying illustration. Corrosion resisting pipe and fittings are also produced by filament winding.

EPSOMITE. Epsomite is normally found as an efflorescence on mine and cave walls. It belongs to the orthorhombic crystal system, being a hydrous sulfate of magnesium, $MgSO_4 \cdot 7H_2O$, with vitreous to earthy luster and colorless to white, of transparent/translucent quality. Hardness of 2–2.5, and specific gravity of 1.68. Very bitter to the taste. Found with the soluble salt lake deposits in Stassfurt, Germany, and in limestone caves in Kentucky, Tennessee and Indiana; also in several California and Colorado abandoned mines.

EQUAL-AREA MAP. A map so drawn that a square mile in one portion of the map is equal in size to a square mile in any other portion. This is obtained by changing the scales along the meridians and parallels in inverse proportion to each other.

Equal-area maps covering the whole globe, such as the Sanson-Flamsteed sinusoidal projection, are approximately elliptical. In Lambert's azimuthal equal-area projection, the parallels of latitude come closer together as the equator is approached (see **Lambert Projection**). Any map from the center of the whole world or of a hemisphere necessarily distorts the shape of a region far from the center of the map, but such maps are useful for some climatological studies in which the correct representation of area is important. For limited parts of the globe, equal-area projections are quite practical. (Cf. **Conformal Map.** See **Map Projections.**)

EQUAL-ENERGY SOURCE. A light source for which the time rate of emission of energy per unit of wavelength is constant throughout the spectrum.

EQUALIZER (Network). It is very difficult to construct apparatus and particularly lines for communication circuits which will pass all the necessary frequencies equally. Often it is necessary to insert networks whose function is merely to restore the original relation of the various frequencies. These are called equalizers. Usually they are loss circuits which will cause more loss in those frequencies which had been attenuated less and the opposite for those attenuated the most. Thus the equalizer shows a frequency response which is the inverse of the system it is intended to equalize, so the result of connecting it in the circuit is to restore the overall response to a flat characteristic. Equalizers may be used to correct either frequency distortion, phase distortion, or both. The term equalizer is also used for the series of connections sometimes made in dc machines to insure equal current distribution in the conductors.

EQUATION. An equality involving two or more numbers or functions. Equations are classified by terms describing the functions in them or they are given the name of a mathematician who discovered or studied the equation. Depending on the functions, the equation could be algebraic or transcendental, with many subclasses in both of these types.

A plane curve is usually described by a single equation in two variables representing rectangular or polar coordinates. Sometimes it is preferable to represent the curve by parametric equations expressing the coordinates separately in terms of a third variable, the parameter. Parametric equations of surfaces and of curves in space are also useful.

The equation of a given locus is the equation which is satisfied by the coordinates of all points of this locus, and of only such points.

When it is of interest to find the solution of an equation, approximate methods are often useful.

EQUATION OF STATE. Also called *characteristic equation*, a relation, empirical or derived, between thermodynamic properties of a substance or system. The equation of state must be single-valued in terms of its variables. This is a direct consequence of the concept of state.

There exist systems, namely systems which undergo processes involving hysteresis (plastic deformation or ferromagnetism, for example) for which no equation of state can be indicated. Although the laws of thermodynamics may apply to such systems, the rigorous results of classical thermodynamics are not applicable because the science of thermodynamics is developed on the assumption of the existence of the single-valued function.

In the realm of classical thermodynamics, equations of state are assumed given. They can be derived from first principles only by the methods of statistical mechanics and quantum mechanics. These rely on the adoption of suitable molecular models for substances, and so far no universal, generally applicable model has been discovered even for narrow classes of substances such as gases.

It is an experimental fact that every thermodynamic system possesses a definite number n of independent properties which determine its state. Consequently, an equation of state is a relation between n properties (mutually independent) chosen (otherwise arbitrarily) as the independent properties ($x_1, x_2 \ldots x_n$) of the system and one more property, the dependent property y. Hence the equation of state is a function of the form

$$F(y, x_1, x_2, \ldots x_n) = 0$$

The simplest thermodynamic systems possess two independent properties, consequently the simplest equation of state is written in terms of three variables. When it is written in terms of pressure p, volume V, and absolute temperature T, it is called the p-V-T relation for the system or the thermal equation of state. When one of the caloric-thermodynamic properties (better called caloric properties, because p, V, T are thermodynamic properties also), such as enthalpy, entropy, Gibbs function, or work function (Helmholtz function) are given, the equation is called a thermodynamic equation, or better, a caloric equation of state, although the latter is not a commonly accepted designation.

Even in the case of a simple system, one equation of state, e.g., the equation $f(p, V, T) = 0$, does not necessarily determine the form of all the other equations of state. This is connected with the fact that the derivation of the other equations of state may involve the integration of partial derivatives which leads to the appearance of whole functions in the integration "constant." An equation from which other equations of state can be derived by differentiation only, is called a *fundamental equation* (of state). In the case of a pure substance in a specified phase, the p, V, T relation does not constitute a fundamental equation with respect to the properties U, H, S, G, A, or their derived properties C_p, C_v, γ, etc. Consequently, it is possible to have two or more substances whose p, V, T relations are identical but whose specific heats, for example, are different.

In the case of continuous systems, for which the state changes from point to point, for example, a flow field of a viscous fluid, it is assumed that at every point, the equation of state is the same as for a homogeneous system and does not involve the gradients of the thermodynamic properties. Hence, such systems can only be studied with the aid of thermodynamics if local departures from equilibrium are small (near-equilibrium processes), i.e., if the gradients of the thermodynamic properties are not too great.

An equation of state must necessarily involve a finite (even if very large) number of independent variables. The particular variables which are chosen as independent is immaterial, on condition that they are mutually independent, and that their number is appropriate to the physical nature of the system.

Equations of states of various types of systems are numerous. The Curie equation is the equation of state of a paramagnetic solid. The Beattie and Bridgeman equation, Berthelot equation, Clausius equation, Dieterice equation, Keyes equation, and van der Waals equation are other examples in this category.

It should be noted that equations of state for systems which consist of several components, rather than a single substance, can be written by introducing the variables $N_1, N_2, \ldots N_c$, which are the respective mole numbers of the components present.

EQUATION OF TIME. The interval by which the true Sun is ahead or behind the mean Sun. The Sun's apparent motion in the sky varies throughout the year because the earth's speed in its elliptical orbit varies slightly from perihelion to aphelion, while the mean Sun is assumed (by definition) to travel with a constant speed equal to the average speed of the true Sun. The interval never exceeds 17 minutes.

EQUATOR. A plane perpendicular to the axis of rotation of the earth and passing through the center of the earth will intersect both the surface of the earth and the celestial sphere in great circles. These great circles are known as the terrestrial and celestial equators. Some navigators refer to the celestial equator as the equinoctial. See also **Thermal Equator.**

EQUATORIAL COORDINATES (Astronomy). Equatorial coordinates are a system of spherical coordinates, in which the origin may be the eye of the observer (in which we have the apparent system of coordinates), the center of the earth (geocentric system), the center of the sun (heliocentric system), or the center of the Milky Way (galactocentric system). The fundamental line in the heliocentric system is the line joining the poles of rotation of the earth, which cuts the celestial sphere in its poles of rotation. The plane perpendicular to the fundamental line through the origin is the celestial equator. The fundamental direction in the plane may be either the point of intersection of the local meridian with the celestial equator, which is above the horizon, or the vernal equinox.

To locate an object in this system of coordinates, a plane is passed through the object and the line joining the poles of rotation, and this plane cuts out a great circle, known as an hour circle, on the celestial sphere perpendicular to the plane of the equator. The declination of an object is the angular distance of the object north (+) or south (−) of the celestial equator measured in the plane of the hour circle through the object. The hour angle of the object is the angular distance, measured in the plane of the equator, from the point of intersection of the meridian above the horizon to the point of intersection of the hour circle through the object, in the direction of apparent rotation (west) of the celestial sphere. The right ascension of the object is the angular distance, measured in the plane of the equator from the vernal equinox to the point of intersection of the hour circle in a direction (east) contrary to the direction of apparent rotation of the celestial sphere. For convenience, both right ascension and hour angle are frequently expressed in units of hour, minutes and seconds of time, rather than the more common angular notation of degrees, minutes, and seconds of arc.

Due to the fact that the local meridian remains fixed as the celestial sphere apparently rotates, the hour angle of an object is continually changing. Since both the vernal equinox and the hour circle rotate with the celestial sphere, both the right ascension and declination of the object remain fixed as the sphere rotates. However, both right ascension and declination change slowly due to precession and nutation. In tabulating these coordinates in star catalogues, the values are given for the position of the equinox for some particular date, and the corrections necessary to reduce the positions to the present date must be applied.

See also **Celestial Sphere and Astronomical Triangle.**

EQUATORIAL COUNTERCURRENT. A current moving eastwards and caused by the return of lighter water piled up on the inner margins by the North Equatorial Current and South Equatorial Current on the western side of the ocean basin. This countercurrent lies in a band of calm air, the doldrums (see **Winds and Air Movement**).

EQUATORIAL TELESCOPE. A telescope so mounted that it may be moved parallel to the equatorial coordinates of hour angle and declination; sometimes called simply an equatorial. In this form of mounting, one axis, known as the polar axis, is parallel to the axis of rotation of the earth, and the other axis, known as the declination axis about which the telescope may be rotated, is perpendicular to the polar axis. From this arrangement of the axes, the telescope as it is rotated about the declination axis, must move in a plane perpendicular to the equator, and hence, parallel to hour circle or in the direction of declination. The rotation of the declination axis about the polar axis, with the telescope remaining fixed in declination, will cause the telescope to move parallel to the equator, and hence, in the coordinate of hour angle.

The majority of equatorials are carried on a single pier. The difficulty with this form of mounting is that the telescope will frequently run into the pier when the hour angle is close to zero. To avoid this, several other methods of supporting the polar axis have been devised. Perhaps the most common is the so-called English mounting, in which the two ends of the polar axis are supported on separate piers, with the telescope free to pass through zero hour angle.

The most difficult adjustment of the equatorial is to get the polar axis strictly parallel to the axis of rotation of the earth. When this has been accomplished, the instrument is by far the most convenient of all forms of mounting. If the instrument is rotated about the polar axis from east to west at exactly the same rate that the earth is rotating about its axis from west to east, the telescope lens will remain fixed relative to objects on the celestial sphere. Hence, once the instrument is set on a star, clockwork may be devised to keep the telescope "following."

See also **Equatorial Coordinates;** and **Telescope.**

EQUILIBRIUM. In the elementary sense of the macroscopic (visible to the naked eye) system, equilibrium is obtained if the system does not tend to undergo any further change of its own accord.

Mechanical and Electromagnetic Systems. Equilibrium in mechanical and/or electromagnetic systems is reached when the vectorial summation of generalized forces applied to the system is equal to zero. In any potential field, that is, gravitational or electric vector potential, force can be expressed as gradient of potential (magnetic force however, is a curl of a vector potential). The potential energy therefore has an extremum at the equilibrium configuration. For example, a system such as a mass suspended by a string against the gravitational force (or its weight) is at mechanical equilibrium if the tensile force in the string is equal to the weight of the mass it supports. The d'Alembert principle further states that the condition for equilibrium of a system is that the virtual work of the applied forces vanishes.

Thermodynamic Systems. When a hot body and a cold body are brought into physical contact, they tend to achieve the same warmth after a long time. These two bodies are then said to be at thermal equilibrium with each other. The zeroth law of thermodynamics (R. H. Fowler) states that two bodies individually at equilibrium with a third are at equilibrium with each other. This led to the comparison of the states of thermal equilibrium of two bodies in terms of a third body called a thermometer. The temperature scale is a measure of state of thermal equilibrium, and two systems at thermal equilibrium must have the same temperature.

Generalization of equilibrium consideration by the second law of thermodynamics specifies that the state of thermodynamic equilibrium of a system is characterized by the attainment of the maximum of its entropy. Thermodynamic coordinates are defined in terms of equilibrium states.

Equilibrium between two phases of a system is reached when there is no net transfer of mass or energy between the phases. Phase equilibrium is determined by the equality of the Gibbs functions (also called free enthalpy, free energy, or chemical potential) of the phases in addition to equality of their temperatures and stresses (such as pressure and/or field intensities—intensive properties). Equilibrium of first-order phase change requires continuity of slope or first derivative of the Gibbs function with respect to an intensive property and is generalized as the Clapeyron relation. Second- and higher-order phase changes are given by the condition of continuity of curvature or second derivative of the Gibbs function and so on.

Chemical or nuclear equilibrium of a reactive system is reached when there is no net transfer of mass and/or energy between the components of a system. At chemical or nuclear equilibrium, the Gibbs function of the reactants and the products must be equal according to stoichiometric proportions, in addition to uniformity in temperature and stresses. Chemical equilibrium is summarized in the form of the Law of Mass Action. The trend for the displacement from an equilibrium state is specified by LeChâtelier's principle.

Thermodynamic equilibrium is reached when the condition of mechanical, electromagnetic, thermal, phase, and chemical and nuclear equilibrium is reached.

Stability of Equilibrium. A process or change of state carried out on a system such that it is always near a state of equilibrium is called a quasi-stationary equilibrium. This requires that the process be carried out slowly. If a mechanical system is initially at the equilibrium position with zero initial velocity, then the system will continue at equilibrium indefinitely. An equilibrium position is said to be stable if a small disturbance of the system from equilibrium results only in small, bounded motion about the rest position. The equilibrium is unstable if an infinitesimal displacement produces unbounded motion. In the gravitational field, a marble at rest in the bottom of a bowl is in stable equilibrium, but an egg standing on its end is in unstable equilibrium. When motion can occur about an equilibrium position without disturbing the equilibrium, the system is in neutral (or labile, or indifferent) equilibrium, an example being a marble resting on a perfectly flat plane normal to the direction of gravity. It is readily seen that stable equilibrium is the case when the extremum of potential is a minimum.

When dealing with general thermodynamic systems, the fact that entropy tends to a maximum in the trend toward equilibrium of a natural process generalizes the above mechanical consideration with respect to stability. An equilibrium state can be characterized as a stable equilibrium when the entropy is a maximum; neutral equilibrium when displacement from one equilibrium state to another does not involve changing entropy; and unstable equilibrium when entropy is a minimum. Any slight disturbance from an unstable equilibrium state of a system will lead to transition to another state of equilibrium.

Statistical Equilibrium. In the microscopic sense, that is, treating systems in terms of elemental particles such as molecules, atoms, and other material or quasi-particles (such as photons in radiation, phonons in solids and liquids), equilibrium states are recognized as the most probable states. An equilibrium state of a system is therefore defined in terms of most probable distributions of its elements among microscopic states which may be defined in terms of energy states. In this sense, statistical equilibrium is a condition for macroscopic equilibrium and an equilibrium state of a system is one of its extremal states. In the methods of statistical mechanics, the probability of distribution is expressed in terms of the density of distributions in the phase space. Based on the Liouville theorem, if a system is in statistical equilibrium, the number of the elements in a given state must be constant in time; which is to say that the density of distribution at a given location in phase space does not change with time. For an isolated system, the distribution is represented by a microcanonical ensemble. At equilibrium, no phase point can cross over a surface of constant energy, and the density of distribution is preserved. In this case individual molecules of a system can be represented by phase points. Any part of an isolated system in statistical equilibrium can be represented by a canonical ensemble. A subsystem of a large system in thermal equilibrium also behaves like the average system of a canonical ensemble. A system and a constant temperature bath together can be considered as an isolated system. A phase point in a canonical ensemble can represent a large number of molecules, thus accounting for strong interactions. A canonical ensemble is characterized by its temperature and is therefore pertinent to the concept of thermal equilibrium. When applied to equilibrium of systems involving mass exchange, such as a chemical system, we have a "particle bath" in addition to a constant temperature bath. The pertinent representation for equilibrium including mass exchange as well as energy exchange is known as a grand canonical ensemble, which accounts for the chemical potentials of its elements.

When applied to a system with a large number of elements, the distributions are measured by thermodynamic probability (W); the most probable distribution is such that W is a maximum. This optimal principle is consistent with the condition of maximum entropy (S) cited under **Entropy.** The Boltzmann hypothesis states that $S = k \ln W$, where k is the Boltzmann constant.

Depending on the specifications of W, namely, those of Maxwell-Boltzmann (for low concentration of distinguishable particles, weak interaction and high temperature, such as a dilute perfect gas), Fermi-Dirac (for elemental particles with antisymmetric wave functions at high concentrations of indistinguishable particles and low temperatures, such as electrons in metal), or Einstein-Bose (for elemental particles with symmetric wave functions, such as He^4 at high concentration of indistinguishable particles and low temperature), equilibrium distributions take different forms. The Maxwellian speed distribution in a dilute perfect gas is a distribution based on Maxwell-Boltzmann statistics.

As a consequence of molecular considerations, when two systems are connected for transfer of mass without significant transfer of energy, such as two containers at different temperatures connected by a capillary tube, we have the relation of thermal transpiration.

Trend toward Equilibrium. The mechanism by which equilibrium is attained can only be visualized in terms of microscopic theories. In the kinetic sense, equilibrium is reached in a gas when collisions among molecules redistribute the velocities (or kinetic energies) of each molecule until a Maxwellian distribution is reached for the whole bulk. In the case of the trend toward equilibrium for two solid bodies brought into physical contact, we visualize the transfer of energy by means of free electrons and phonons (lattice vibrations).

The Boltzmann H-theorem generalizes the condition that with a state of a system represented by its distribution function f, a quantity H, defined as the statistical average of $\ln f$, approaches a minimum when equilibrium is reached. This conforms with the Boltzmann hypothesis of distribution in the above in that $S = -kH$ accounts for equilibrium as a consequence of collisions which change the distribution toward that of equilibrium conditions.

Consideration of perturbation from an equilibrium state leads to methods for dealing with rate processes and methods of irreversible thermodynamics in general.

Fluctuation from Equilibrium. A necessary consequence of the random nature of elemental particles in a body is that the property of such a body is not at every instant equal to its average value but fluctuates about this average. A precise meaning of equilibrium can only be attained from consideration of the nature of such fluctuations. In the above, we have repeatedly considered a "large" number of particles. It is important to know how large a number is "large." When considering fluctuation of energy from an average value in an isolated system, the ratio of the two is given to be proportional to $1/\sqrt{N}$, where N is the total number of elements in the system. This is also the magnitude of the fluctuation of number of particles in a system involving transformation of phases and chemical and nuclear species. An equilibrium state is one at which the longtime mean magnitude of fluctuation from the average state is independent of time and this magnitude has reached a minimum value.

Large perturbation from a given state of fluctuation leads to a relaxation process toward a state of equilibrium. The relaxation time, for instance, measures the deviation from quasistationary equilibrium of a process which is carried out at a finite rate.

EQUILIBRIUM DIAGRAM. A diagram showing the phase fields of an alloy system under the conditions of complete equilibrium using as coordinates the temperature, the compositions in terms of the components, and the pressure. The most frequently used equilibrium diagrams in metallurgy are drawn with the pressure considered constant. See iron-carbon diagram under **Iron Metals, Alloys, and Steels.** See also **Distillation.**

EQUINES. See **Horses, Asses, and Zebras.**

EQUINOX. The line of intersection of the plane of the earth's equator with the plane of the ecliptic (the line of nodes of the earth) intersects the celestial sphere in two diametrically opposite points known as the equinoxes. As seen from the earth, the sun apparently passes through each of the equinoxes once each year, passing through the vernal equinox on approximately March 21st and through the autumnal equinox on approximately September 22nd.

The great circle passing about the celestial sphere through the equinoxes and the pole of the ecliptic is known as the *equinoctial colure.*

The *autumnal equinox*, for either hemisphere, is the equinox at which the sun "retreats" into the opposite hemisphere. In northern latitudes, the time of this occurrence is approximately September 22nd. The *vernal equinox*, for either hemisphere, is the equinox at which the sun approaches from the opposite hemisphere. In northern latitudes, this occurs approximately on March 21st. See also **Celestial Sphere and Astronomical Triangle.**

EQUIVALENCE PRINCIPLE. It is always possible at a point in space-time to transform to a (in general accelerated) coordinate system such that the effects of gravity will disappear over a differential region in the neighborhood of the point. As a particular case, if there are two observers, one uniformly accelerated with acceleration g and not in a gravitational field, the other not accelerated but held in a uniform gravitational field g, the results of mechanical and optical experiments performed by the two observers will be identical. See **Gravitation;** and **Relativity and Relativity Theory.**

EQUIVALENCE THEOREM. The field in a source-free region bounded by a surface could be produced by a distribution of electric and magnetic currents on that surface that would be equivalent, for points inside the surface, to the actual external sources.

EQUIVALENT CIRCUIT. This term is applied to an electrical circuit which is electrically equivalent to another circuit, or sometimes, to a mechanical device. Equivalent circuits of mechanical systems or electromechanical systems such as loudspeakers enable the designer to apply methods of circuit analysis and often obtain a solution easily which would be very difficult if not impossible otherwise. The equivalent circuit method is used extensively in the analysis of communication circuits, particularly those involving vacuum tubes and transistors. These circuits do not yield the static currents and voltages (those existing in the absence of applied signals) but their use provides a method of computing the response to alternating voltages or dc changes, provided both are of small magnitude. Both of these are usually the ones of interest. Similarly, many other types of electrical circuits may be simplified in terms of equivalent circuits, sometimes giving all the necessary solutions, sometimes giving solutions for limited conditions, but, in most cases, greatly decreasing the labor involved in analyzing the circuit or equipment.

EQUIVALENT ELECTRONS. For an atom, electrons in the same orbital (whereby they have the same principal quantum number and the same azimuthal quantum number). For a molecule, electrons having the same quantum numbers, apart from spin, and the same symmetry g or u.

EQUIVOCATION. A measure of the average ambiguity of a received signal. It is the residual remaining uncertainty when the received signal has been interpreted.

ERA (Geology). See **Geologic Time Scale.**

ERBIUM.[1] Chemical element symbol Er, at. no. 68, at. wt. 167.26, eleventh in the Lanthanide Series in the periodic table, mp 1529°C, bp 2868°C, density 9.066 g/cm^3 (20°C). Elemental erbium has a close-packed hexagonal crystal structure at 25°C. The pure metallic erbium is silver-gray in color and retains its luster at room temperature, not affected by moisture or normal atmospheric gases. Large pieces of the metal do not oxidize readily even when heated. Fine chips and powder, however, will ignite and burn. Because of its comparative softness, the metal can be worked by conventional equipment. The metal should be annealed after size-reduction. There are six natural isotopes ^{162}Er, ^{164}Er, ^{166}Er through ^{168}Er and ^{170}Er. Twelve artificial isotopes have been pre-

[1]The main portion of this article was revised and updated by K. A. Gschneidner, Jr., Director, and B. Evans, Assistant Chemist, Rare-Earth Information Center, Energy and Mineral Resources Research Institute, Iowa State Univ., Ames, Iowa.

pared. The natural isotopes are not radioactive. In terms of abundance, erbium is present on the average of 2.8 ppm in the earth's crust, making its potential availability about equal with uranium. The element was first identified by C. G. Mosander in 1843. The thermal-neutron-absorption cross section of erbium is 160 borns per atom, relatively high and tenth among the natural elements. The metal has a low acute-toxicity rating. Electronic configuration

$$1s^22s^22p^63s^23p^63d^{10}4s^24p^64d^{10}4f^{11}5s^25p^65d^16s^2.$$

First ionization potential 6.10 eV; second 11.93 eV. Ionic radius Er^{3+} 0.881 Å. Metallic radius 1.758 Å. Other important physical properties of erbium are given under **Rare-Earth Elements and Metals.**

Erbium occurs in certain types of apatites, xenotime, and gadolinite. These minerals also are processed for their yttrium content as well as for other heavy *Lanthanide* elements. With liquid-liquid organic and solid-resin organic ion-exchange techniques, the separation of erbium from the other elements is favorable.

Because of the metal's high thermal-neutron-absorption cross section, it has been of much interest in terms of use in nuclear reactor hardware. When an erbium-activated phosphor is coated onto a gallium-arsenide diode, the latter emits infrared radiation, which is converted to visible light by the phosphor. Through variation of the energizing power and by use of a combination of rare-earth-activated phosphors, the primary colors of light can be produced. Thus, erbium holds promise for use in display panels and color-television picture tubes. An erbium hydride-hydrogen system at a fixed temperature creates an extreme vacuum and when used for comparative purposes makes it possible to measure vacuums in the range of 10^{-4} to 10^{-11} torr with much precision. The system has been used for the calibration of ionization gauges used for very high vacuums as found in outer space. Erbium is in an early stage of investigation for application to lasers, semiconductor devices, garnet microwave devices, ferrite bubble devices, and catalysts.

See references listed at ends of entries on **Chemical Elements;** and **Rare-Earth Elements and Metals.**

Erbium in Lightwave Communication Amplifier. In January 1992, E. Desurvire (Columbia University Center for Telecommunications Research) reported that optical fibers made from silica glass and traces of erbium can amplify light signals when they are energized by infrared radiation. Desurvire developed an efficient radiation source (referred to as a laser diode chip) that, when integrated into a fiber optic communication system, can increase transmission capacity by a factor of 100.

The device can be effective in very long stretches of communication cable, such as used in transoceanic service. Each cable will be capable of carrying 500,000 messages simultaneously, a factor of 12 times greater than present cables.

ERG. See **Units and Standards.**

ERGODICITY. Generally, this word denotes a property of certain systems which develop through time according to probabilistic laws. Under certain circumstances a system will tend in probability to a limiting form which is independent of the initial position from which it started. This is the *ergodic property*. A stationary stochastic process (x_t) may be regarded as the set of all realizations possible under the process. Each such realization may have a mean m_r. If the process itself has a mean $E(x_t) = \mu$, the ergodic theorem of Birkhoff and Khintchine states that m_r exists for almost all realizations. If, in addition, $m_r = \mu$ for almost all realizations the process is said to be ergodic. In this sense ergodicity may be regarded as a form of the law of large numbers applied to stationary processes.

ERGOT. A fungus which grows upon and replaces the grain of rye. Ergot in the human body contracts smaller arteries and smooth muscle fibers, especially that of the uterus. Ergotism is a diseased condition of humans and domestic animals resulting from eating grasses or grain infected with ergot fungus; or from excessive uses of the drug made from the fungus. Ergotamine, one of the alkaloids of ergot, has been used in the treatment of migraine.

ERIDANUS. A very long constellation that extends from the equator to the southern horizon. The constellation contains the star Achernar.

ERMINE. See **Mustelines.**

EROSION (Geology). A fundamental process of geology which brings about alterations in rocks and other major features of the physical earth. Some of the results of erosion are beneficial, as exemplified by the gradual disintegration, transformation, and transportation of certain kinds of rocks to form soil. On the other hand, erosive processes can leach out or carry away excellent soil, and thus destroy agricultural productivity, sometimes over wide areas, by the action of wind and water. For soil erosion, see **Soil.**

In a general sense, erosion refers to the reduction of the land surface toward sea level by the various agencies of weathering, stream action, glacial action, wind action, chemical action, and other forces. A majority of these forces occur naturally and essentially are beyond control by way of human intervention. On the other hand, erosion also results from unplanned (in terms of long-term) erosion by various projects which may create or accelerate already existing natural erosive processes. Alteration of the course of streams and certain types of agricultural methods can cause serious erosion problems.

The surface of the earth, subject to both natural and artificial forces, is constantly undergoing an overall process that may be termed *weathering*. Erosion is a major result of this process. Erosion processes may be classified into two major groups: (1) *chemical erosion*; and (2) *mechanical erosion*. The actions of erosive agents, such as air, wind, and water, may act both in a chemical and mechanical fashion. See also **Glacier.**

Chemical Erosion

The principal agents of chemical erosion are water and air. Both agents attack the content of the original rocks and thus bring about changes in chemical composition with accompanying physical changes. In falling through the atmosphere, rain water (*meteoric water*) picks up small amounts of oxygen, carbon dioxide, and other gases (notably in areas where air pollutants are present). The addition of these ingredients to otherwise pure water increases the solvent power of the water. Water in soaking downward through the soil reaches rocks which often contain numerous cracks and fissures, thus affecting the rock to cause a considerable area of exposed surface. Further, the solvent power of the water may be increased as it acquires new chemical substances from the rocks themselves. For example, water attacking pyrite will pick up sulfur and through a complex process ultimately convert this to sulfuric acid, which is highly corrosive. Particularly, the combined action of oxygen in the air with moisture from rain can become highly corrosive, as witness the immediate formation of rust upon exposed iron (ferrous) materials. Water plus carbon dioxide can yield corrosive carbonic acid. Thus, in processes of oxidation and carbonation, the chemical weathering of rocks in both erosive and corrosive. Water alone through the process of hydration can cause decomposition of some rocks. For example, two parts of hematite, Fe_2O_3, will unite with three parts of water to form limonite (iron rust), $2Fe_2O_3 \cdot 3H_2O$.

The erosive results of groundwater are principally chemical in nature. Although the corrosive solutions may move slowly, over periods of time extensive changes in rock structures can be brought about, particularly the creation of a large thickness of mantle rock. Thus, enormous caverns may be created, such as the Carlsbad Cave (New Mexico) or the Mammoth Cave (Kentucky). Also, rich ore deposits may result from such actions, as witness the great copper deposits found in the western United States or the iron deposits of the Lake Superior region. The analysis of ground waters testifies to their prior erosive performance. In descending order of content, groundwaters will be found to contain (1) calcium carbonate and calcium sulfate; (2) colloidal silica; (3) sodium carbonate, sodium sulfate, and sodium chloride; (4) magnesium carbonate; and (5) potassium carbonate—all materials derived from prior contacts with rocks and soil. To a lesser extent, ocean water also works in a similar fashion along the coasts.

The mechanical effects of groundwater become significant when the volume and velocity of the water reach a point where the flow can pick up and transport solid particles of rock. Sometimes this process follows or is concurrent with the aforementioned chemical actions of groundwater. The lower portion of Mammoth Cave, exhibiting gravels, sands, and muds, provides evidence of the extensive action of groundwater. The principal results of the mechanical action of groundwater which may be observed from the surface include soil creep, landslides, and rock streams. The underlying slippery nature of shales and clays when wet assist the process of alteration. There are rock streams in the Rocky Mountains of the United States a mile or more in length. So-called "walking mountains" (as reported in China) are the result of such phenomena on an extremely large scale. See **Cave.**

The mechanical actions of the oceans are far more important than their chemical actions in terms of erosion. These include the actions of waves, currents, and tides. A *terrace* or *wave-cut beach* is formed by the action of waves cutting into a land surface of moderate relief. A *sea-cliff* may result where relatively high land is undercut by waves. The chalk cliffs of England and France are examples. Where particularly soft, vulnerable rock structures are encountered by wave action, *sea caves* may be formed. The Blue Grotto (Naples) is a notable example. Sometimes spouting caves are formed. See **Blowhole.** Where rocks contain vertical joints, these may be eroded at a greater rate, producing a cliff with deep indentations. Where the eroded rock may become separated from the land mass, it may form a *chimney* or *stack*. See also **Chimney Rock.** Where there is a rock connection remaining at a higher elevation, a *natural bridge* will be the result.

Waves of the oceans also may form deposits along the shore, built from the deposition of materials by the mechanical work of the waves. Such terms as *wave-built terraces*, *barrier beaches*, *spits*, *hooks*, *bars*, and *tombolos* are used to describe these resulting features. See **Barrier Beach.**

The wind also serves as a mechanical agent for picking up, carrying, and later depositing solid materials. Numerous erosive features result. Wind-worn pebbles (glyptoliths) give evidence of wind erosion. These are also referred to as dreikanters. See **Dreikanter.** The ravages of wind removal and transportation of soil are only too evident in certain dust bowl areas, generally the result of poor agricultural methods.

When material is moved by wind and ultimately released, two dominant types of deposits occur: (1) *loess*; and (2) *dunes*. See **Dune.** Loess is composed of dust and silt, usually comprised of quartz grains, ranging in size from about 0.1 millimeter, and clay. Loess is characterized by a buff-to-yellow color and great porosity. Loess may be observed along both sides of the Mississippi River (Louisiana and Mississippi, north Illinois and Iowa); also the Missouri River and other tributaries in the Mississippi Valley. Loess occurs extensively in central Europe and in Tibet, Mongolia, and China. Thicknesses of loess deposits in China up to 300 feet (91 meters) are reported. The deposits in Europe are usually quite thin; those in the United States range from 10 to 20 feet (3 to 6 meters) in thickness.

The particle size of sand in sand dunes ranges in diameter from 0.05 millimeter up to that of coarse gravel. Most dunes are made up of quartz, but gypsum dunes are found in New Mexico. Dunes of calcareous oolites are found in the Bahamas and Bermuda; rather low dunes of dry clay occur in Montana.

Although of a seasonal nature, snow drifts are an aspect of the dune phenomenon.

Dunes range from a few feet to over 400 feet (120 meters) in height. They may cover a few square feet (a fraction of a square meter) to up to several square miles (kilometers). The dunes in desert regions commonly are 200 to 400 feet (60 to 122 meters) in height. Usually the axis of a dune is at right angles to the prevailing wind. Nearly all sand dunes migrate to a measurable extent unless they are covered with vegetation. Migrating dunes will advance upon forests, farms, highways. Dunes buried a number of cities of ancient times (Babylonian, Chaldean). Migration of dunes can be slowed or stopped through the planting of surface grasses and shrubs of a type that will withstand the prevailing climate.

The carrying of volcanic dust is a special form of wind deposit. Such deposits may occur long distances from their sources, as witness the volcanic dust deposits in southern Nebraska and north-central Kansas.

Cycle of Erosion

This is a term used to generally describe the work of rivers and streams—erosional, transportational, and depositional. The erosional work of rivers and streams sculptures the surface of the earth into a variety of forms. The river erosion pattern of any region will depend upon the climate, the relative hardness and solubility of the formations, the structure, and the degree to which the erosive process has completed its work, with or without interruptions caused by diastrophism. The stream pattern of a region is not only indicative of the structural control, but also of the stage in its erosional history. The ideal complete cycle of erosion begins with uplift of a region with low altitude and ends with reduction of the uplifted region to a peneplain. See accompanying diagram.

Successive stages in the normal cycle of erosion in a region of folded rocks.

Soil Conservation through Erosion Control

Modern programs of erosion control have been organized on a large scale in the United States, Australia, New Zealand, and in several countries in Africa, south of the Sahara. In nearly all countries where agriculture is important there has been recognition of the problem of erosion.

The cultivation of irrigated rice in oriental countries reduces erosion to a minimum on nearly all of the land actually used for rice. In China and India, much of the land between the irrigated paddy fields has lost its upper layers of soil to sheet erosion and is scored by deep gullies. This kind of problem has been alleviated in Japan by intensive reforestation and by other erosion-control measures on land used for unirrigated crops.

Erosion in many European countries has been kept at a low level for several generations through enlightened land-use programs and intensive cultivation and fertilization; but much of the upland soil bordering the Mediterranean Sea has been lost to erosion. Fertile floodplains around the Mediterranean have been badly damaged by deposition of coarse debris washed from cultivated uplands.

Modern soil conservationists recommend using the land in ways that will produce the greatest income consistent with the least loss of soil. They endeavor to protect sloping lands by planting soil-conserving crops, by contour cultivation, by alternation of strips of close-growing crops with clean-cultivated ones, and by using graded terraces and grass-covered waterways to control the rate of flow or runoff of water. Shallow gullies may be filled by plowing, and deep gullies may be controlled by check dams or vegetative cover. However, terraces and dams may be harmful if not well maintained. The large volume of water collected behind terraces and check dams sometimes breaks through, sweeping away large volumes of soil, and scours out new gullies. Ditches that were dug to drain wet lowland soils for farming have provided lower base levels for water flow, and as a consequence gullies have eaten their way back from the ditches high into the adjacent uplands. Many examples of this exist in eastern Nebraska and western Iowa.

The bad effects of wind erosion can be ameliorated on cultivated land by planting strips of close-growing crops in alternation with strips of sod or fallow land arranged at right angles to the prevailing wind. Stubble mulch in semiarid wheat land is highly effective. List-furrowing across the wind is effective in some areas, and shelterbelts of trees and shrubs have been helpful to a limited extent. Wind erosion of mulch and peat beds can be greatly reduced by planting wind-breaks at right angles to the prevailing wind.

EROSION (Soil). See **Soil.**

ERRATIC. In geology, an ice-carried boulder or block, sometimes weighing many tons, which because of its lack of similarity to the bedrock or formation on which it rests, and the peculiarity of its position, must have been transported to its present resting place by a glacier or an iceberg. When erratics occur in sufficient quantity to form relatively pronounced topographic features these are called moraines. When erratics of similar or identical lithology show a well-defined lineal distribution from the parent outcrop they are called boulder trains.

ERROR. The word error may be used in two ways: 1. In general, it is a mistake in the colloquial sense, for example, an error in copying, an error of reference, or an error of interpretation. 2. In statistics, the word error is used to denote the difference between an occurring value and its "true" or "expected" value. Specific types of errors and error functions follow:

Error, Approximation. In general, an error due to approximation in numerical calculations as distinct; e.g., an error of observation. More particularly, a rounding error.

Error, Experimental. In general, any error in an experiment whether due to stochastic variation or bias. It is the aim of good experimental design to provide valid measures of the experimental error in the more restricted sense.

Error Band. In estimation or prediction, the estimated or predicted value is bracketed by a range of values (determined by standard errors, confidence-intervals or similar methods) within which the value may be supposed to lie with a certain probability. This is called the error band.

Error Function. The definite integral, also called the Gauss error function,

$$\operatorname{erf}(t) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-y^2} dy$$

When the results of a series of measurements are described about an average by a Gaussian curve, erf(ha) is the probability that the error of a single measurement lies between $\pm a$, where h is the precision index.

Values of the integral, as functions of t, have been tabulated. Sometimes the more general function

$$E_n(t) = n! \int_0^t e^{-y^n} dy$$

is discussed and $E_2(t)$ is called the error integral.

Error in Equations. An equation in variables or variates may be inexact, either because the equation is not a complete representation of the situation (as in a demand-supply equation which omits other factors such as income or employment) or because it is disturbed by extraneous sources of variation (as in an autoregression equation). These departures from the relationship expressed by the equation are known as errors in the equation; as distinct from effects such as observational errors in the variables themselves.

Error of Estimation. In general, the difference between an estimated value and the true value. More specifically, in regression analysis where the regression equation is used to estimate the "dependent" from given values of the "independent" variates, the difference between the estimated and the observed value of the dependent variate.

Error of First Kind. If, as the result of a statistical test, a statistical hypothesis is rejected when it ought to be accepted, i.e., when it is true, then an error is committed. This class of error is termed an error of the first kind and is fundamental to the theory of testing statistical hypotheses associated with the names of Neyman and (E.S.) Pearson. The frequency of errors of the first kind can be controlled by an appropriate selection of the regions of acceptance and rejection; that is to say, by choice of appropriate critical regions it is possible to ensure that the probability of committing an error of the first kind is an assignable constant.

Error of Measurement. Much of the routine work in any experimental research in the physical sciences is concerned with the eliminating, minimizing, or compensating for observational errors. By the error of any measurement is meant the result of the individual measurement minus the true value of the quantity measured. It may thus be either positive or negative. Errors of measurement may be broadly classified into two types: instrumental errors and personal errors. An instrumental error is any error in measurement which results from the properties of the instruments used in the measurement. Instrumental errors may be divided into scale errors, which result from improper calibration of the instrument, and reproducibility errors, which result from the failure of the instrument to give the same indication whenever it is subject to the same input signal. The latter type may be treated as accidental errors, the former may not. A personal error is any error which results from the tendency of an observer to misread an instrument, e.g., to read consistently high or low. Personal errors are not distributed in the same manner as accidental errors, and their magnitudes may be estimated only by the comparison of observations made by different observers. An error of sampling is the discrepancy between the estimate derived from a sample and the true value. The occurrence of errors in this sense is inherent in the incompleteness of sample coverage; they are not "mistakes" in the ordinary sense of the word.

Error of Observation. An error arising from imperfections in the method of observing a quantity, whether due to instrumental or to human factors.

Error of Second Kind. If, as the result of a test, a statistical hypothesis is accepted when it is false, i.e., when it should have been rejected, then an error has been made. This class of error is termed an error of the second kind and, like errors of the first kind it is fundamental to the Neyman-Pearson theory of testing statistical hypotheses. Unlike the error of the first kind, however, it is not, in general, controlled by the simple process of selecting regions of acceptance and rejection. The customary procedure in choosing tests of hypotheses is to fix the magnitude of the first kind of error and, with this restriction, to minimize the second kind of error.

Error Signal. See **Signal (Instrument).**

Error Variance. The variance of an error component. Thus, if the generating model of a set of data consists of certain systematic components together with a stochastic component, the variance of the latter is in the error variance. The expression can also be understood in a wider sense, as the variance of error in repetitions of an experimental situation, whether the "error" is due to sampling effects or not. It makes for clarity if expressions such as "error variance" are eschewed in favor of "residual variance" but the use of the former type of wording is very widespread.

Sir Maurice Kendall, International Statistical Institute, London.

ERROR (Common-Mode). See **Common-Mode Voltage.**

ERUPTIVE. This term has the same general geological meaning as effusive, but is sometimes used in the much more general sense as synonymous with igneous.

ERYSIPELAS. An acute inflammation of the skin (*superficial cellulitis*) due almost always to infection by Group A streptococci. In the newborn, Group B streptococci may cause the disease. The lesion is red, indurated, edematous, and spreads peripherally. Usually the bridge of the nose and cheeks are involved, but other areas of the face and head (ears) may be the site of the lesion. Fever is present from the onset. The disease in most cases is self-limiting. With the use of penicillin therapy, improvement may be achieved within one or two days, but the lesion may not subside for several days.

ERYTHRITE. A mineral of the composition, $Co_3(AsO_4)_2 \cdot 8H_2O$, isomorphous with annabergite. The color ranges from rose to crimson. The mineral sometimes contains nickel and occurs in monoclinic crystals, in earthy forms (as a weathering product of cobalt ores) in the oxidized portions of the veins, or in globular and reniform masses. The mineral sometimes is referred to as *erythrine, cobalt bloom*, and *peachblossom ore*.

ESCARPMENT. A term used by physiographers and structural geologists to denote a line of steep slopes or cliffs.

ESKER. Certain long, often winding, ridges formed of stratified sands and gravels, which occur within the glaciated regions of Europe and North America are called eskers. These pronounced topographic features are frequently several miles in length and because of their peculiar and uniform shape somewhat resemble railway embankments. Eskers represent the deposits of glacial streams which flowed within and under glaciers. After the retaining ice walls melted away the stream deposits remained as long winding ridges. Synonyms for esker include os, eschar, serpent kame, and Indian ridge.

ESOPHAGUS. The tubelike passage that connects the lower end of the pharynx (throat) with the stomach. This tube is about 10 inches (25 centimeters) long. It lies in front of the spine as it descends through the chest and passes through the diaphragm just before it enters the stomach. The walls of the esophagus are made up of circular and straight muscle fibers which allow for wavelike contractions to move food downward toward the stomach. The inner surface contains many glands which secrete mucus for lubrication of the walls. Difficulty of swallowing is a symptom associated with a number of esophageal disorders and is called *dysphagia*. Diagnostic aids sometimes used in identifying esophageal disorders include barium contrast x-rays, esophagoscopy (using a fiberoptic instrument which permits examinations of the tube), and an acid perfusion test, the latter being particularly helpful in the diagnosis of esophagitis. A weak hydrochloric acid solution is administered through a nasoesophageal tube alternately with a saline solution. The presence of pain in response to hydrochloric acid during the test cycle is an excellent indicator of esophagitis and helps the physician differentiate this condition from the similar pain of coronary artery disease.

The principal disorders of the esophagus are as follows. (1) *Chronic peptic esophagitis* is quite common and usually responds well to intermittent antacid therapy. Onset of the condition usually commences an hour or so after meals. There may be regurgitation of small amounts of gastric contents into the mouth. The typical symptom is commonly referred to as *heartburn*. Unresponsive cases may require surgery. This condition is not always associated with a hiatus hernia. (2) *Hiatus hernia* is caused by a portion of the stomach protruding through the hiatus of the diaphragm and into the thoracic cavity. Barium x-ray examination and esophagoscopy are usually required to affirm the disorder. Because the symptoms of chronic peptic esophagitis and hiatus hernia are so similar, a physician may commence with antacid therapy. If the condition persists, cimetidine therapy may be initiated. In the absence of success, surgical repair of the hernia is indicated, along with restoration of the gastroesophageal junction. (3) *Carcinoma of the esophagus*, to date, has proved to be a very high risk situation. Because of late diagnosis, only about 20% of patients are surgical candidates. The surgery is technically difficult, and may be accompanied by radiation therapy. The five-year survival rate after surgical therapy alone has been estimated at about 10%. Dysphagia is an early symptom of this condition. Other disorders of the esophagus include: (4) lower esophageal ring; (5) diffuse esophageal spasm; (6) achalasia; (7) collagen vascular disease; and (8) esophageal diverticulum.

See also **Digestive System (Human).**

ESSENTIAL AMINO ACIDS. See **Amino Acids.**

ESSENTIAL HYPERTENSION. See **Hypertension (High Blood Pressure).**

ESTERIFICATION. See **Cellulose Ester Plastics (Organic); Esters.**

ESTERS. The compound resulting from the reaction of an alcohol with an acid is termed an *ester*. The reaction is termed *esterification* and is accompanied by the yield of H_2O along with the ester. The reaction is highly reversible and hydrolysis will occur in the reverse direction when H_2O remains present. The formation of ethyl nitrate from ethyl alcohol and nitric acid typifies a simple esterification: $C_2H_5OH + HNO_3 \rightleftharpoons C_2H_5NO_3 + H_2O$.

Under normal conditions, esterification occurs slowly and inasmuch as the reaction is fully reversible, an equilibrium is reached which tends to withhold completion of the reaction in either direction.

The esterification reaction can be speeded up by the use of a catalyst. Such a catalyst is hydrogen ion, as from HCl or H_2SO_4. Side reactions may occur, HCl furnishing some organic chloride, and H_2SO_4 causing dehydration of the alcohol. Phosphoric acid generally avoids both these results. Salts that hydrolyze to furnish hydrogen ions are also used, e.g., zinc chloride, aluminum sulfate, ferric chloride, sometimes by the addition of acid to these salts.

The equilibrium point can be displaced to produce more ester by increasing the relative amounts of either the acid or the alcohol as desired.

A complicating factor of considerable significance when recovery of the ester is to be made by distillation is the existence of 2-component (binary) azeotropes of constant boiling points with (1) ester and H_2O and (2) ester and alcohol, and 3-component (ternary) azeotropes with (3) ester and H_2O and alcohol. See article on **Azeotropic System**.

Esters are high-tonnage chemicals. Among the more important esters are normal, secondary, and isobutyl acetates, ethyl acetate, normal, secondary, and isoamyl acetates, and methyl acetate. These acetates are used primarily in the lacquer industry. Cellulose nitrate is used in the plastics, lacquer, and explosives industries; cellulose acetate in the plastics and lacquer industries; glyceryl trinitrate in the explosives industry; while cellulose xanthate (viscose) is an important synthetic product for textiles. In the specialized field of plasticizers for the plastics and lacquer industries, numerous synthetic esters are used; as examples, butyl stearate, diamyl phthalate, dibutyl oxalate, dibutyl phthalate (also for smokeless powder), dibutyl sebacate, dibutyl tartrate, diethylene glycol monostearate, diethylene glycol distearate, diethyl phthalate, dimethyl phthalate, diphenyl phthalate, glyceryl tripropionate, isobutyl phthalate, tributyl borate, tributyl citrate, tributyl phosphate, tricresyl phosphate, triethylene glycol dihexoate, triethylene glycol dioctoate, triethyl citrate, triethyl phosphate, triphenyl phosphate. Methyl methacrylate ester is an important plastic. Ethyl silicate is used to cover concrete, brick and stone with a coating of silicic acid to resist water penetration. Diocytl phthalate is used as a plasticizer in cable and wire insulation, and dimethyl phthalate as an insect repellent.

See also **Organic Chemistry.**

ESTIMATION (Theory of). Given a sample from a population whose specification involves one or more parameters, it is necessary to form estimates of the parameters. Usually, many different estimates of a given parameter can be derived, and the theory of estimation is concerned with the properties of these different estimates. Together with the estimate itself, it is useful to provide some idea of its precision and this is commonly done by specifying an interval which is intended to contain the true value of the parameter (see **Confidence Interval;** and **Fiducial Inference**).

A basic result in the theory of estimation states that, under general conditions, a consistent estimator has a sampling variance in large samples which is not less than a certain lower bound. Statistics whose variance attains this lower bound are said to be efficient. A still more valuable property, which applies for all sample sizes, is that of sufficiency; a sufficient statistic is one which summarizes all the information on the parameter that is contained in the sample. R. A. Fisher has shown that the method of maximum likelihood leads to efficient estimators, and to sufficient estimators where these latter exist. The method of minimum chi-square also leads to efficient estimates in large samples, but in finite samples may fail due to the occurrence of small expected frequencies.

ESTROGEN. See **Gonads; Hormones; Steroids.**

ESTROGEN (Infertility). See **Infertility.**

ESTUARY. The wide mouth of a river, or arm of the sea, where the tide meets the river current, or flows and ebbs. It may also be defined as "a body of water in which the river water mixes with and measurably dilutes seawater" (Ketchum, 1951). These definitions do not overlap completely, because a lagoon connected with the sea may also be affected by the tide.

Some scientists prefer to describe the environment in terms of the salinity of the water (saline, brackish, or fresh), but saline water is not restricted to marginal marine areas. Such a description does not consider, therefore, the most characteristic aspect of the estuarine environment—that it is a region of steep and variable gradients in the environmental conditions (Fig. 1).

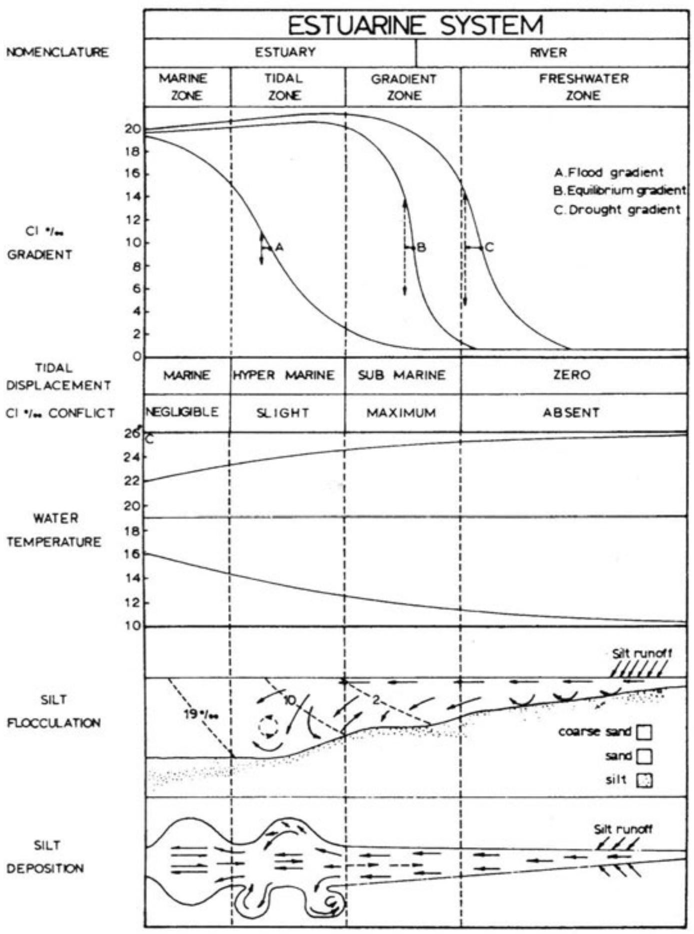


Fig. 1. Some zonal features of a composite Australian estuarine system.

If the physiography of estuaries is a sole consideration, they can be defined as "bodies of water bordered by and partly cut off from the ocean by land masses that were originally shaped by nonmarine agencies. They are usually perpendicular to the coast line and most of them occupy the drowned mouths of stream valleys and are, therefore, usually considered as evidence of submergence" (Emery and Stevenson, 1957).

Classification of Estuaries. There is no system which is universally used to classify estuaries. A broad classification separates normal (or positive) estuaries, in which freshwater inflow exceeds evaporation, from inverse (hypersaline) estuaries in which evaporation exceeds freshwater inflow. Neutral estuaries are those in which neither evaporation nor river discharge dominates.

Estuaries along most coastlines have been formed partly by the subsidence of the land mass and partly by the rise in sea level. These embayments are usually elongate indentures of the coastline with rivers flowing in from the landward ends. Deep estuaries are known as *rias*. In eastern North America most estuaries are shallow with irregular, or dendritic, shore lines and are normal estuaries.

Along the Gulf Coast of the United States, marine processes have built a series of barrier islands parallel to the coastline. Most of the islands extend across the mouths of estuaries, forming a lagoon and decreasing the width of the estuarine entrance to the open sea.

The exchange of water, in such cases, between the estuary and the open sea is modified by the intervening lagoon in which evaporation may exceed freshwater inflow. The waters in the estuary, then, have salinities higher than normal as a result of the exchange with the lagoonal water.

Water Characteristics and Circulation. The important feature in an estuary is the intermixing of seawater with the freshwater from land drainage. This interaction usually produces a variation, both horizontal and vertical, in the salinity of estuarine waters. In normal estuaries, salinities range from nearly zero at the river's mouth, to approximately 30% at the seaward extremity. In addition, there is generally an increase in salinity with depth.

An inverse estuary also has greater salinities at depth, but the highest salinities are at the head of the embayment rather than at the mouth. There may be a difference of several parts per thousand between the salinity at the head and that of normal seawater (Fig. 2).

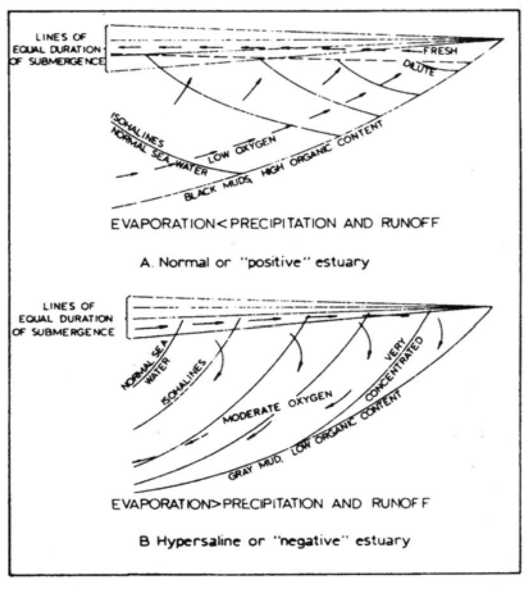


Fig. 2. Schematic sections of the two basic types of estuaries.

Temperature. The water in estuaries, and especially that overlying the tidal flats, is relatively thin, so it follows the variations in temperature of the atmosphere more closely than does the water of the open sea. The water is much colder in winter and warmer in summer than is the sea. The diurnal variation is also greater than in the sea.

There are pronounced variations in water temperature with depth. During the winter, the water is cold and nearly isothermal at all depths. In some instances, in response to cold weather, the surface water may become a degree or so colder than the deeper water. During the summer, solar radiation and minor wind mixing produce a high temperature at the surface with less change at depth. The difference between surface and bottom temperatures may also be influenced by warm or cold river water flowing over the dense seawater.

Where evaporation exceeds river inflow, the summer surface water may become so saline as to sink to the floor of the estuary and flow out of the entrance beneath the incoming seawater. The temperature of the deep water may then be higher than that at the surface.

Circulation. Water movements in estuaries result mainly from the interaction of tides, river flow, and wind. The tides and river flow are usually the dominant factors. In Gulf Coast estuaries, where the tide range is small and river discharge is at times negligible, wind-induced currents are most important.

Stable estuaries. In normal estuaries, the distribution of temperature and salinity, and the circulation pattern, are controlled almost exclusively by the tide range and the river inflow. Tidal currents tend to produce turbulent mixing of the river water and the seawater. However, the low-density freshwater above the seawater results in a stable vertical stratification which resists mixing. As a consequence, the relative magnitudes of the river discharge and the tidal flow are significant in controlling the physical structure of the water in the estuary.

Where the river flow is large in relation to the tidal exchange, the seawater enters the estuary as a salt-water "wedge" along the bottom. However, there is frictional drag between the overlying freshwater, the salt-water wedge, and the bottom. The relative velocities of the seaward-flowing freshwater and the intruding salt-water wedge control the magnitude of the friction factor. Thus, the actual position of the wedge is closely dependent on the volume of the river flow. When the volume of river discharge is great, the wedge extends only a short distance into the estuary and, of course, vice versa.

The salinity of the salt-water wedge remains similar to that of the open sea because there is only minor mixing with the seaward-flowing freshwater. However, at the interface between the two types of water, waves form and sometimes intrude into the surface water. Thus, the salt content in the upper layers increases slightly as the water moves seaward. Even so, throughout the estuary, a sharp salinity gradient exists between the two water layers.

The loss of salt water from the wedge to the upper layer is compensated by a flow of water from the sea (Fig. 2). The exchange from below takes place all along the upper interface of the wedge. As a result, there is a flow directed upstream at all positions within the wedge. The landward-moving water in the salt wedge is minor, however, and of the two, the seaward flow of surface water is far greater.

Partly mixed estuaries. Where tidal movements are great as compared to the volume of river discharge, mixing between the seawater and freshwater is sufficient to destroy sharp interfaces. The salt wedge, in such cases, does not exist as an identifiable feature, but a transition layer of definitely increasing salinity does occur. In such an estuary, however, the salinity in both the upper and lower layers decreases toward the head of the estuary.

The chief cause of currents in estuaries in which the waters are partly mixed is the tide. As in the stratified estuary, there is a net water movement superimposed on the tidal currents—a net seaward flow at the surface and a net flow toward the head in the deeper layers. These water motions are not as well defined as in a stratified estuary, and there is no sharp current interface. The flow from the deeper layers toward the surface decreases toward the head of the embayment (Fig. 2). The volume rate of seaward flow increases, therefore, toward the mouth.

Mixed estuaries. In wholly mixed estuaries, the movements induced by the tide are far greater than those produced by the river inflow. The waters are completely mixed and are isohaline from the surface to the bottom. At all depths, the salinity decreases from the mouth to the head.

In such estuaries, the outward flowing water is deflected to the right, in the northern hemisphere, because of earth rotation. Thus, in wide estuaries, the salinity is less on the right side (looking toward the estuarine mouth) than on the left. A net seaward flow exists along the right side and a net landward flow on the left. Water also moves laterally across the estuary from the left to the right side resulting in horizontal mixing.

In narrow, well-mixed estuaries, mixing induced by tidal action may be great enough to eliminate any lateral salinity gradient. There is a net seaward flow in all waters and the only difference in salinity is the normal decrease toward the head.

Estuaries bordered by lagoons. Along coastal regions where barrier islands extend across the mouths of estuaries, the water bodies are usually so shallow that mixing by winds is sufficient to produce homogeneous water. Tidal currents are only significant through the inlets between the barrier islands. The total volume of water which flows in and out is relatively small. As a consequence, the rise and fall of the tide and tidal currents are minor within the estuary, and the most significant currents are from wind action.

There is, necessarily, a net flow of water out of these shallow estuaries sufficient to remove the water added by freshwater discharge. The large cross-sectional area of the estuary and the dampening effect of the coastal lagoon reduce the flow so that it is normally not directly measurable. The net motion may be completely modified by wind action to the extent that high water and constant, net inflow may occur during times when prevailing winds blow up the estuary. Strong winds blowing from the land reverse this effect and result in extremely low water levels in the estuary and the extrusion of estuarine waters many miles to sea.

Seiches in estuaries. In some of the larger estuaries, the periodic flooding by the tides is supplemental by seiches, long stationary waves. The simplest seiche is one whose node is at the mouth of the estuary and the antinode near the head. The period of the seiche is controlled by the length and depth of the body of water, and where its natural period nearly coincides with that of the tide, as at the Bay of Fundy, a great fluctuation in sea level occurs (about 15 meters). In most estuaries, the seiche is only a few centimeters and is obscured by the much greater tidal amplitude.

Waves. Wind waves are small in estuaries because of the short fetch and the shallow water. They usually cause little erosion although when the tide is high, waves may stir up the muddy sediment on tidal flats. Waves may transport some sand and, because the largest waves come across the widest part of the estuary, they form sand spits pointing upstream in tidal channels.

A tidal bore (a wave of translation) is common in narrow estuaries and tidal channels. As a result of the shape of the entrance and bottom friction, the flooding tide is held back for a time until the water finally rushes up the channel as a steep wall of water. Bores may be from a few centimeters to several meters in height and move at velocities as great as 10 knots. The character of a tidal bore is determined, in part, by the river discharge which must be sufficient to hold back the tide for a period of time.

Estuarine sediments. The inorganic sediments of estuaries are derived from inflowing rivers bordering sea cliffs, the sea floor outside the estuary, and the reworked deposits of tidal flats and marshes along the shores. Regardless of the source, much reworking of sediment occurs within estuaries. Erosion, too, is evident from the migration of tidal channels and the muddy color of the water when no river inflow is taking place. Some estuaries have entrances narrow enough so that tidal currents scour the bottom locally, leaving rocky or gravelly bottoms. The prevailing condition, however, must be one of deposition, and the average rate of deposition is greater than that of the open sea.

Additional Reading

Armstrong, R.: "White Oak River Dying," *Tideland News*, Swansboro, North Carolina, February 20, 1980.

Baumann, R. H., Day, J. W., Jr., and C. A. Miller: "Mississippi Deltaic Wetland Survival: Sedimentation Versus Coastal Submergence," *Science*, **224**, 1093-1095 (1984).

Clark, J., and S. McCreary: "Prospects for Coastal Resource Conservation in the 1980s," *Oceanus*, **23**, 4, 22-31 (1981).

Czaya, E.: "Rivers of the World," Van Nostrand Reinhold, New York, 1982.

Emergy, K. O., and R. E. Stevenson: "Estuaries and Lagoons," *Geol. Soc. Am. Mem.*, **67, 1,** 673-750 (1957).

Hey, R. D., Bathurst, J. C., and C. R. Thorne: "Gravel-Bed Rivers," Wiley, New York, 1982.

Ketchum, B. H.: "The Flushing of Tidal Estuaries," *Sewage Ind. Wastes*, **23,** 2, 198-209 (1951).

Lauff, G. H. (editor): "Estuaries," American Association for the Advancement of Science, Publ. 83, Washington, D.C., 1967.

Lynch, D. K.: "Tidal Bores," *Sci. Amer.*, **247**(4), 145–156 (October 1982).

McLusky, D. S.: "The Estuarine Ecosystem," Wiley, New York, 1981.

Nichols, F. H., et al.: "The Modification of an Estuary," *Science*, **231**, 567–573 (1986).

Pillsbury, A. F.: "The Salinity of Rivers," *Sci. Amer.*, **245**(1), 54–65 (July 1981).

Salati, E., and P. B. Vose: "Amazon Basin: A System in Equilibrium," *Science*, **225**, 129–138 (1984).

Staff: "Tender Sediments," *Sci. Amer.*, **254**(1), 68 (January 1986).

ETHANE. C_2H_6, formula weight 30.07, colorless, odorless gas, mp $-172°C$, bp 88.6°C, sp gr 1.05 (air = 1.0), practically insoluble in H_2O, moderately soluble in alcohol. The compound burns when ignited in air with a pale faintly luminous flame; forms an explosive mixture with air over a moderate range. With excess air, products of combustion are CO_2 and H_2O. Ethane is among the chemically less reactive organic substances. However, ethane reacts with chlorine and bromine to form substitution compounds. Ethane occurs, usually in small amounts, in natural gas. The fuel value of ethane is high, 1,730 Btu per cubic foot. Ethane may be prepared by reaction of magnesium ethyl iodide in anhydrous ether (Grignard's reagent) with H_2O or alcohols. Ethyl iodide, bromide, or chloride are preferably made by reaction with ethyl alcohol and the appropriate phosphorus halide. Important ethane derivatives, by successive oxidation, are ethyl alcohol, acetaldehyde, and acetic acid.

ETHANOL. See **Ethyl Alcohol.**

ETHANOLAMINES. There are three ethanolamines, all hydroxyamines, and all high-tonnage industrial chemicals. Production is about 300 mil pounds annually.

Monoethanolamine, $NH_2CH_2CH_2OH$, industrial symbol (MEA) Formula weight 61.08, mp 10.5°C, bp 171°C, sp gr 1.018
Diethanolamine, $NH(CH_2CH_2OH)_2$, industrial symbol (DEA) Formula weight 105.14, mp 28.0°C, bp 270°C, sp gr 1.019
Triethanolamine, $N(CH_2CH_2OH)_3$, industrial symbol (TEA) Formula weight 149.19, mp 21.2°C, bp 360°C, sp gr 1.126

Mono- and triethanolamine are miscible with H_2O or alcohol in all proportions and are only slightly soluble in ether. Dietanolamine will dissolve in H_2O up to 96.4% at 20°C, is very soluble in alcohol, and only slightly soluble in ether.

Wurtz first reported the ethanolamines in 1860, but they were not used commercially on any scale until the late 1920s. All of the compounds are clear, viscous liquids at standard conditions and white crystalline solids when frozen. They have a relatively low toxicity. Industrially, the compounds are important (1) because they form numerous derivatives, notably with fatty acids, soaps, esters, amides, and esteramides; and (2) for their exceptional ability for scrubbing acidic compounds out of gases. Monoethanolamine, for example, will effectively remove H_2S from hydrocarbon gases. The compounds also remove CO_2 from process streams and, where desired, the CO_2 may easily be recovered by heating the absorptive solutions. The soaps of the ethanolamines are extensively used in textile treating agents, in shampoos, and emulsifiers. The fatty acid amides of diethanolamine are applied as builders in heavy-duty detergents, particularly those in which alkylaryl sulfonates are the surfactant ingredients. The use of triethanolamine in photographic developing baths promotes fine grain structure in the film when developed. Ethanolamine also is used as a humectant and plasticizing agent for textiles, glues, and leather coatings; and as a softening agent for numerous materials. Morpholine is an important derivative.

In early processes, the ethanolamines were prepared by reacting ethylene chlorohydrin $ClCH_2 \cdot CH_2OH$ with NH_3. Current processes react ethylene oxide $\langle(CH_2)_2\rangle O$ with NH_3, usually in aqueous solution. The ratio of mono-, di-, and triethanolamines varies in accordance with the amount of NH_3 present. This is controlled by the quantities of MEA and DEA recycled. Higher NH_3-ethylene oxide ratios favor high DEA and TEA yields, whereas lower ratios are used where maximum production of MEA is desired. The reaction is noncatalytic. The pressure is moderate, just sufficient to prevent vaporization of components in the reactor. The bulk of the H_2O produced in the reaction is removed by subsequent evaporation. The dehydrated ethanolamines then proceed to a further drying column, after which they are separated in a series of fractionating columns, not difficult because of the comparatively wide separation of their boiling points.

e (The Number). A transcendental number, used as the base of the system of natural or Napierian logarithms. It is defined by

$$e = \lim_{n\to\infty}(1 + 1/n)^n$$

or by

$$e = \lim_{x\to 0}(1 + x)^{1/x}$$

It is represented by the infinite series

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \cdots + \frac{1}{n!} + \cdots$$

and it equals 2.71828, approximately. In this book, the number *e* is written in italic form, except in equations involving electrical quantities, where the Roman e is used to avoid confusion.

See also **Logarithm.**

ETHERS. The homologous series of ethers has the formula $C_nH_{2n+2}O$. Structurally, the ethers have an oxygen linkage between two radicals (R—O—R′). R and R′ may be the same as in dimethyl ether $CH_3—O—CH_3$; or they may differ as in ethylisopropyl ether $C_2H_5—O—C_3H_7$. The latter may be referred to as a *mixed ether*. Mixed ethers frequently are made from mixed alcohols. Where R and R′ are alkyls, the ether may be called an *alkyl ether* or an *alphyl oxide*. They may be considered to be derivatives of the monohydric alcohols. Each ether is isomeric with a saturated alcohol. Both diethyl ether and butyl alcohol are $C_4H_{10}O$. Also, there are many isomeric ethers, starting with $C_4H_{10}O$. Methylpropyl ether and diethyl ether are isomeric. Where compounds such as these have the same general formula, are members of the same family, and differ only by the alkyl group present, they are termed *metameric*.

Since they are similar structurally to the alcohols, phenols also form ethers. An example of an aromatic ether is methylphenyl ether (anisole) $C_6H_5—O—CH_3$. There are few ethers where both R and R′ are aryls. The structure of thioethers is similar to the other ethers, but with a sulfur atom in the link instead of an oxygen atom, as R—S—R′. Examples of thioethers include diethyl sulfide $C_2H_5—S—C_2H_5$ and methylethyl sulfide $CH_3—S—C_2H_5$, which is a mixed thioether.

The properties of the ethers may be summarized by: (1) with the exception of dimethyl ether which is a gas, the ethers are volatile, mobile, inflammable liquids that are lighter than H_2O; (2) they are relatively inert chemically, not being acted on by alkali metals or alkalis and not reacting with dilute acids; (3) they form substitution products when reacted with chlorine and bromine; and (4) they are decomposed when heated with strong acids, yielding esters.

Ether. $(C_2H_5)_2O$, formula weight 74.12, mp $-116.3°C$, bp 34.6°C, sp gr 0.708. Probably the best known of the ethers, diethyl ether, commonly called simply *ether*, is slightly soluble in H_2O (1 volume in 10 volumes H_2O) and is miscible with alcohol in all proportions. Ether dissolves iodine and many organic substances, e.g., oils and fats, waxes, resins, and alkaloids and hence is widely used as a solvent for these substances in the preparation of numerous products, including explosives and collodion. Ether explodes in oxygen in the presence of a flame or spark, yielding H_2O and CO_2. When heated with an acid, such as H_2SO_4, ether yields ethyl alcohol. With phosphorus halides, ethyl halide (2 moles) is formed. Ether reacts with HNO_3 to form ethyl oxide.

Although still used medically, at one time ether was the major anesthetic, for which it must be scrupulously pure. In addition to various side effects which may result from the use of ether as an anesthetic, it is a definite hazard in the operating room because of its explosive properties, particularly in enriched oxygen atmospheres.

See also **Organic Chemistry.**

ETHOLOGY. The study of animal behavior, particularly under natural conditions. Embraces the concepts of altruism and other aspects of sociobiology.

ETHYL ALCOHOL. C_2H_5OH, formula weight 46.07, colorless liquid with mild characteristic odor, mp $-$ 114.1°C, bp 78.32°C, sp gr 0.789. Also known as *ethanol*, the compound is miscible in all proportions with H_2O or ether. When ignited, ethyl alcohol burns in air with a pale blue, transparent flame, producing H_2O and CO_2. The vapor forms an explosive mixture with air and is used in some internal combustion engines under compression as a fuel. See also **Fuel.** Such mixtures are frequently referred to as *gasohol.*

Anhydrous ethyl alcohol is made from the constant boiling mixture with H_2O (95.6% ethyl alcohol by weight)—(1) by heating with a substance such as calcium oxide, which reacts with H_2O and not with alcohol, and then distilling, or (2) by distilling with a volatile liquid, such as benzene (bp 79.6°C) which forms a constant low-boiling mixture with H_2O and alcohol (bp 64.9°C), so that H_2O is removed from the main portion of the alcohol; after which alcohol plus benzene distills over (bp 78.5°C). Anhydrous ethyl alcohol is required for certain purposes as a solvent and reagent and fuel applications.

Commercially, ethyl alcohol is marketed by the proof gallon, 200 proof on the scale representing pure alcohol (100%). When the term *alcohol* alone is used, it refers to a liquid that ranges from 188 to 192 proof (94% to 96% ethyl alcohol). When the terms *grain alcohol, high-purity* alcohol, or *pure* ethyl alcohol are used, these usually refer to a liquid that is 190 proof. In most countries, beverage alcohol is highly taxed and to make the product available for nonbeverage purposes, denaturants will be added. Denaturants include methyl alcohol, pyridine, benzene, kerosene, pine oil, mixtures of primary and secondary aliphatic higher alcohols, and hydrogenated organic compounds. Thousands of nonbeverage industrial and commercial products, notably food extracts, toiletries, pharmaceuticals, solvents, and cleaning products, contain denatured ethyl alcohol.

Worldwide, ethyl alcohol is the basis for a huge alcoholic beverage industry, offering a wide range of products wherein the alcoholic content varies from a few to over 50% (100 proof). Industrially, ethyl alcohol is very important high-tonnage raw and intermediate material for numerous processes, and is used extensively in solvents, antiseptics, antifreeze compounds, and fuels.

Production. Natural fermentation is the oldest process for making ethyl alcohol and still constitutes the principal means for creating the alcoholic content of beverages. Except in connection with other alcohol-containing products, industrial producers of ethyl alcohol use processes other than fermentation. For fermentation, almost any agricultural raw material with a carbohydrate content in the form of sugars or starches that are easily converted to sugars can be used. Once the raw materials are in the form of sugars, yeast enzymes are added to commence natural fermentation. Traditionally, in the United States, industrial alcohol prepared by fermentation has used blackstrap molasses, which contains up to 50% sugars and can be easily fermented. The starting mash is prepared by diluting the molasses with H_2O to bring the sugar content down to about 15% (weight). The mash is slightly acidified, after which invertase (enzyme to convert sucrose) and zymase (enzyme to convert glucose and fructose) are added. The products are ethyl alcohol and CO_2. Yeast activity is sustained by the addition of nutrients. With careful control of temperature and acidity, the fermentation process can be completed in about two days. The resulting mash (beer) usually contains about 12% ethyl alcohol which is recovered from the beer by distillation. See **Fermentation.**

In modern industrial ethyl alcohol plants, the compound is produced in two principal ways: (1) by *direct hydration of ethylene*, or (2) by *indirect hydration of ethylene*. In the direct hydration process, H_2O is added to ethylene in the vapor phase in the presence of a catalyst: $CH_2{:}CH_2 + H_2O \rightleftarrows CH_3CH_2OH$. A supported acid catalyst usually is used. Important factors affecting the conversion include temperature, pressure, the $H_2O/CH_2{:}CH_2$ ratio, and the purity of the ethylene. Further, some by-products are formed by other reactions taking place, a primary side reaction being the dehydration of ethyl alcohol into diethyl ether: $2C_2H_5OH \rightleftarrows (C_2H_5)_2O + H_2O$. To overcome these problems, a large recycle volume of unconverted ethylene usually is required. The process usually consists of a reaction section in which crude ethyl alcohol is formed, a purification section with a product of 95% (volume) ethyl alcohol, and a dehydration section which produces high-purity ethyl alcohol free of H_2O. For many industrial uses, the 95%-purity product from the purification section suffices.

In the indirect hydration process, ethylene first is absorbed in concentrated H_2SO_4 to form mono- and diethyl sulfates: $CH_2{:}CH_2 + H_2SO_4 \rightarrow CH_3CH_2OSO_3H$; and

$$2CH_2{:}CH_2 + H_2SO_4 \rightarrow (CH_3CH_2)_2SO_4.$$

The ethyl sulfates then are hydrolyzed to ethyl alcohol: $CH_3CH_2OSO_3H + H_2O \rightarrow CH_3CH_2OH + H_2SO_4$; and

$$(CH_3CH_2)_2SO_4 + 2H_2O \rightarrow 2CH_3CH_2OH + H_2SO_4.$$

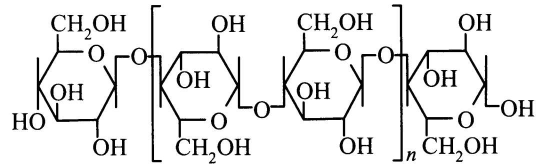
Remaining steps in the process include recovery and purification of the crude ethyl alcohol and reconcentration of the dilute H_2SO_4. The crude ethyl alcohol is steam-stripped from the dilute acid solution, followed by distillation for purification.

Azeotropes. The physical properties of ethyl alcohol are influenced by the hydroxyl group that imparts hydrogen-bonding characteristics and polarity to the substance that are analogous to water. Ethyl alcohol displays a highly nonideal behavior in numerous solutions, forming several azeotropes. The list of binary azeotropes of ethanol is long, including acetonitrile, benzene, carbon disulfide, chloroform, ethyl acetate, hexane, toluene, and water. See also **Azeotropic System.**

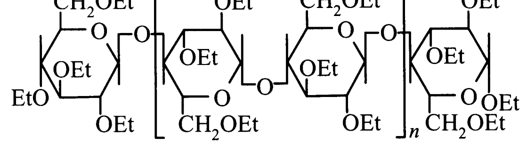
Chemistry. Ethyl alcohol reacts (1) with sodium metal, forming sodium ethoxide C_2H_5ONa plus hydrogen gas, (2) with phosphorus chloride, bromide, iodide, forming ethyl chloride, bromide, iodide, respectively, (3) with H_2SO_4 concentrated, forming at 100°C ethyl hydrogen sulfate $C_2H_5OSO_2OH$, at 140°C diethyl ether $(C_2H_5)_2O$, at 200°C ethylene $CH_2{:}CH_2$, (4) with organic acids, warmed in the presence of H_2SO_4, forming esters, e.g., ethyl acetate $CH_3COOC_2H_5$, ethyl benzoate $C_2H_5COOC_2H_5$ (see various individual acids), (5) with magnesium methyl iodide in anhydrous ether (Grignard's solution), forming methane as in the case of primary alcohols, (6) with calcium chloride to form a solid addition compound $4C_2H_5OH \cdot CaCl_2$, which is decomposed by H_2O, (7) with oxygen, using sodium dichromate solution and H_2SO_4, to form acetaldehyde (and acetic acid), using air, in the presence of acetic bacteria, to form vinegar (dilute acetic acid along with the substances present in the alcohol used, e.g., wine, cider), (8) with HNO_3 (a) concentrated, free from nitrogen tetroxide, to form ethyl nitrate, (b) dilute to form glycollic acid, (c) concentrated acid containing nitrogen tetroxide (fuming HNO_3) explosive reaction, (9) with chlorine (or bromine) to form chloral CCl_3CHO (or bromal).

See **Organic Chemistry.**

ETHYL CELLULOSE. A versatile thermoplastic cellulose ether that is compatible with a wide variety of solvent systems, resins, oils, and plasticizers. This versatility permits a wide diversity of end-product properties. It is an excellent film former as from a wide range of neat, lacquer, or dispersion formulations. Molded ethyl cellulose has excellent toughness, flexibility, and shock resistance. Useful temperature range is from about −40 to +100°C. In the preparation of ethyl cellulose, wood pulp or cotton linters with a high alpha-cellulose content are reacted with ethyl chloride and sodium hydroxide. Structural formulas of cellulose and ethyl cellulose with complete (54.9%) ethoxyl substitution are:



Cellulose, $n > 500$



Tri-*O*-ethyl cellulose, n = 50–150; Et = CH_3CH_2

The natural color of ethyl cellulose is colorless to light amber, but it can be formulated into a wide range of transparent, translucent, and opaque colors. The material should be dried before molding because it is slightly hygroscopic. Compression molding temperatures range from 121–200°C and pressures from 500–5,000 psi. Injection molding temperatures range from 175–260°C and pressures from 8,000–

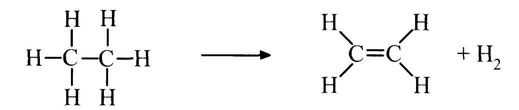
32,000 psi. Strong acids decompose the material, but weak acids and strong alkalis have only a slight effect. Weak alkalis do not attack the material. Ethyl cellulose is soluble in a large number of organic solvents.

Among the commercial uses for ethyl cellulose are strippable coating for metal parts, paper coatings, and in medicinal tablets. An interesting application is coating for bowling pins. Ethyl cellulose sheeting is tough, flexible, and transparent, yet sufficiently rigid to withstand rough handling.

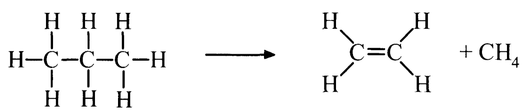
ETHYL CHLORIDE. See **Chlorinated Organics.**

ETHYLENE. C_2H_4, formula weight 28.03, colorless gas with slight odor, normal bp $-103.7°C$, critical pressure of 49.98 atmospheres, and critical temperature of 9.5°C, density 1.26 grams per liter (0°C and 760 mm), sp gr 0.97 (air = 1.0), very slightly soluble in H_2O, slightly soluble in alcohol. Ethylene burns when ignited in air with a luminous flame. The presence of ethylene in coal gas is chiefly responsible for the luminosity of the latter gas. Ethylene forms an explosive mixture with air and has a high fuel value, 1,615 Btu per cubic foot.

Even though there are few direct end-uses foe ethylene, it is probably the most important petrochemical feedstock, both in terms of quantities used and economic value. Ethylene is the feedstock for ethylene oxide, ethylbenzene, ethyl chloride, ethylene dichloride, ethyl alcohol, and polyethylene, most of which, in turn, are used to produce hundreds of other end-products. Most ethylene is produced by steam cracking of ethane or propane.



Ethane Ethylene



Propane Ethylene Methane

Ethylene also may be produced from other paraffinic or naphthenic hydrocarbons. The reactions are highly endothermic (34,400 kcal/kg mole of ethane cracked at approximately 900°C) and proceed in the direction indicated at temperature exceeding approximately 620°C without a catalyst.

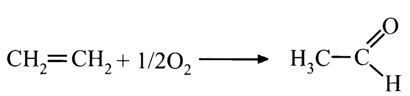
Ethylene is of importance as a petrochemical feedstock because of its great versatility in reacting to form several chemical intermediates. The double bond provides reactivity; the compound also has the ability to homopolymerize and copolymerize with other monomers. Some of the important reactions involving ethylene include:

Chlorination

$$CH_2{=}CH_2 + HCl \xrightarrow[\text{cat.}]{\text{acidic}} CH_3{-}CH_2Cl$$

Ethyl chloride

Oxidation



Acetaldehyde

Hydration

$$CH_2{=}CH_2 + H_2O \longrightarrow C_2H_5OH$$

Ethyl alcohol

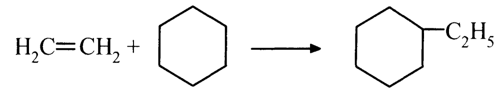
Oxychlorination

$$CH_2{=}CH_2 + 2HCl + 1/2O_2 \longrightarrow C_2H_4Cl_2 + H_2O$$

Ethyl dichloride

Ethylene dichloride is used for the production of vinyl chloride.

Alkylation



Ethyl benzene

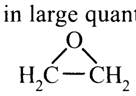
Ethyl benzene is used for the production of styrene.

Polymerization

$$nCH_2{=}CH_2 \longrightarrow (-CH_2-CH_2)_n$$

High- and low-density polyethylene

Ethylene is also oxidized in large quantities to ethylene oxide:



At one time, ethylene was produced by the dehydration of ethyl alcohol over alumina.

Almost any naphthenic or paraffinic hydrocarbon heavier then methane can be steam-cracked to yield ethylene. The preferred feedstock in the United States has been ethane and/or propane recovered from natural gas, or from the volatile fractions of petroleum. However, because of long-term uncertainties pertaining to natural gas, many producers have been turning to heavier petroleum fractions, such as gas oils, as feedstocks. The consumption of ethylene throughout the free world is estimated to be about 40×10^9 pounds per year.

Ethylene reacts (1) with the halogens to form substitution halides; (2) with hypochlorous and hypobromous acid to form ethylene chlorohydrin or ethylene bromohydrin, respectively; (3) with hydrogen iodide or bromide (not chloride) to form ethyl iodide or ethyl bromide; (4) with hydrogen, in the presence of a catalyst, e.g., finely divided nickel at 150°C, to form ethane; (5) with concentrated sulfuric acid at 160°C to form ethyl hydrogen sulfate; and (6) with potassium permanganate to form ethylene glycol, although glycol is preferably made from ethylene dichloride or chlorohydrin.

In addition to its uses in the preparation of intermediates for a large variety of petrochemical reactions, ethylene is used as an anesthetic, as a fuel with oxygen for high-temperature flames, and as a coloring and ripening agent for citrus fruits and tomatoes. Ethylene chlorohydrin is used as an agent for decreasing the dormant period of seeds. See also **Polyethylene.**

ETHYLENE DICHLORIDE. See **Chlorinated Organics.**

ETHYLENE GLYCOL. This compound, $HOCH_2CH_2OH$, is traditionally associated with its use as a permanent-type antifreeze for internal-combustion engine cooling systems. However, since the early-1960s, large tonnages of ethylene glycol have been used in the production of polyesters for fibers, films, and coatings. The compound also finds important uses in hydraulic fluids, in the manufacture of lowfreezing-point explosives, glycol ethers, and deicing solutions. Di-and triethylene glycols are important coproducts usually produced in the manufacture of ethylene glycol. Diethylene glycol, $HOCH_2CH_2OCH_2CH_2OH$, is used in the production of unsaturated polyester resins and polyester polyols for polyurethane-resin manufacture, as well as in the textile industry as a conditioning agent and lubricant for numerous synthetic and natural fibers. It is also used as an extraction solvent in petroleum processing, as a desiccant in natural gas processing, and in the manufacture of some plasticizers and surfactants. Triethylene glycol,

$$HOCH_2CH_2OCH_2CH_2OCH_2CH_2OH,$$

finds principal use in the dehydration of natural gas and as a humectant.

In one process for the manufacture of the aforementioned glycols, ethylene oxide is formed by direct oxidation of ethylene with oxygen over a silver catalyst. After purification, the stabilized ethylene oxide is mixed with a large excess of water, preheated, and fed to an ethylene oxide reactor. Here the ethylene oxide and water react under high temperature and high pressure to form principally ethylene glycol, with the other aforementioned glycols as coproducts. The crude glycols are dehydrated and then recovered individually as highly pure overhead streams from a series of vacuum-operated purification columns.

Principal properties of the glycols are summarized in the accompanying table.

PROPERTIES OF ETHYLENE, DIETHYLENE, AND TRIETHYLENE GLYCOLS

	Ethylene Glycol	Diethylene Glycol	Triethylene Glycol
Molecular weight	62.07	106.12	150.17
Boiling point (760 mm Hg)	197.6°C	245°C	287.4°C
Vapor pressure (20°C)	0.06 mm Hg	<0.01 mm Hg	<0.01 mm Hg
Specific gravity (20°C/20°C)	1.1155	1.1184	1.1254
Freezing point in air (760 mm Hg)	−13°C	83°C	99°C
Water solubility	———————— complete ————————		

SOURCE: Glycols, Shell Chemical Company Bull. IC: 67-58.

ETHYLENE OXIDE. $\langle(CH_2)_2\rangle 0$, formula weight 44.05, liquid, mp −111.3°C, bp 13.5°C, sp gr 0.887. The compound is miscible in all proportions with H_2O or alcohol and is very soluble in ether. Ethylene oxide is slowly decomposed by H_2O at standard conditions, converting into glycol $CH_2OH \cdot CH_2OH$. Ethylene oxide is a very high-tonnage chemical, approaching nearly 4 billion pounds annually. In terms of consumption (1) 60% for manufacture of ethylene glycol, the latter being an antifreeze compound as well as a raw material for production of polyethylene terephthalate used in the manufacture of polyester fibers, (2) 12% for preparation of surfactants, (3) 8% for the manufacture of ethanolamines, (4) 10% for production of ethylene glycols which are used in plasticizers, solvents, and lubricants, and (5) 10% for making glycol ethers which are used as jet-fuel additives and solvents. See also **Rocket Propellants;** and **Antimicrobial Agents (Foods).**

Direct oxidation of ethylene in the presence of a silver catalyst is the predominant large-scale process used: $CH_2{:}CH_2 + \frac{1}{2}O_2 \rightarrow \langle(CH_2)_2\rangle O$. The yield is approximately 70% of the theoretical. For maximum yield, very careful temperature control is required, the yield dropping as the temperature climbs. The side reaction: $CH_2{:}CH_2 \rightarrow 3CO_2 + 2H_2O$ is the main factor for reducing yield. Thus far, silver has proved to be the most effective catalyst. Several compounds have been investigated that can inhibit the side reaction and also be compatible with the catalyst. These compounds have included ethylene dichloride, ethylene dibromide, alcohol, amines, and organometallic compounds, but their success has been limited. Plants have been designed to use either air or pure oxygen for oxidation. Selection presents an interesting study in economics because (1) where air is used, a purge reactor and associated purge absorber are required (not required by the O_2 process), and (2) where O_2 is used, both a CO_2 removal system and an O_2-making facility are required. The trend is toward the oxygen system with the ethylene oxide plant located near an air-separation plant.

Studies that still are inconclusive have linked ethylene oxide with leukemia and stomach cancer. It is estimated that in the United States approximately 270,000 workers are routinely exposed to ethylene oxide. Comparatively high level exposures include 96,000 persons working in hospitals and an additional 21,000 persons who work in commercial medical supply sterilization facilities, as well as in the production of spices and pharmaceutical products. Since the 1950s, ethylene oxide has been used as a sterilizing agent.

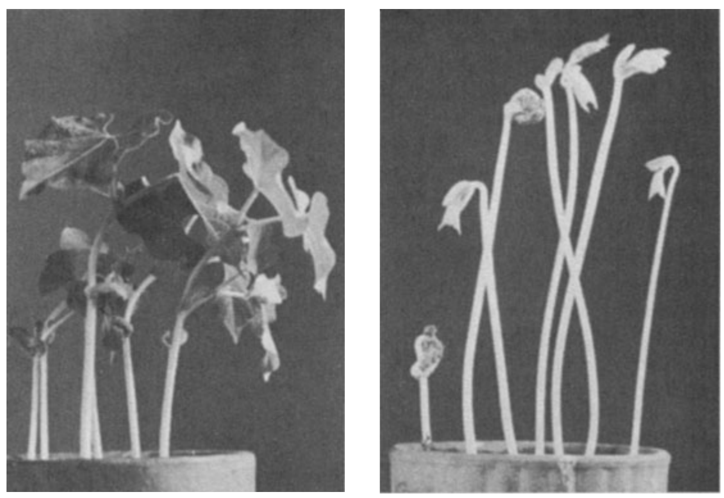
Biochemically, ethylene oxide is a highly reactive epoxide and is a direct alkylating agent. Details of one study are given by K. Steenland, et al. (National Institute for Occupational Safety and the National Cancer Institute) in the *New England J. Medicine*, 1402 (May 16, 1991). The report concludes, "Although our study is the largest to date of workers exposed to ethylene oxide, the results for the relatively rare cancers of a priori interest are still limited by the small numbers of cases and perhaps limited by the short follow-up. Our findings are therefore not conclusive."

It has been established, however, that ethylene oxide is a potent mutagen and animal carcinogen.

ETHYLENE-VINYL-ACETATE COPOLYMERS. Known as EVA copolymers, these materials are polyolefins which can be processed like other thermoplastics, but which approach rubbery materials in softness and elasticity. The resins meet regulatory requirements for use in direct contact with food in food-processing machinery and in packaging applications. They are used in a number of applications to replace plasticized polyvinyl chloride and rubber. EVA copolymers require no curing or plasticizer. Parts made from EVA have little or no odor. Their elasticity is permanent. The copolymers can be injection-, blow-, compression-, transfer- and rotationally molded or extruded into film, sheeting, pipe, and profiles. EVA copolymers offer advantages over polyvinyl chloride and rubber in that they have good clarity and gloss, stress-crack resistance, good barrier properties, low-temperature flexibility and toughness, good adhesive properties, and good resistance to ultraviolet radiation. Their main limitation is a comparatively low resistance to heat and solvents. The resins soften at a temperature of about 70°C. EVA copolymers are not attacked by alcohols, glycols, or weak organic acids. However, to a varying degree, the materials are attacked by chlorinated hydrocarbons, straight-chain paraffinic solvents, and benzene and its derivatives.

ETIOLATION. This is the effect of darkness on a living plant. It is a matter of common observation that plants grown in darkness contain little or no chlorophyll and so are nearly white. Green plants placed in darkness lose their chlorophyll. Eventually, when the food reserves are exhausted, the plant dies.

Besides the lack of chlorophyll, plants grown in the dark have other characteristics. In dicotyledons, the internodes of the stem become excessively elongated and very slender. The leaves fail to expand normally. In monocotyledons, the leaves become very long and usually very narrow, but the stem shows little change. Etiolated plants never bear flowers, unless the flower buds are well developed before the plants are darkened.



Effect of darkness on the elongation of stems. The bean seedlings on the left were grown in the light; those on the right were grown in darkness.

Internally, the tissues are soft and lack strength, the cells being very large and having thin walls. Very little differentiation occurs, the conducting tissues being very much reduced. Leaves which form in darkness show very little of the structure characterizing a normal green leaf, but are almost entirely composed of loosely arranged parenchymatous cells.

See also **Plant Growth Modification and Regulation.**

ETIOLOGY. Knowledge of the cause of any disease or abnormal condition.

EUCALYPTUS TREES. Of the family *Myrtiaceae* (myrtle family), eucalpytus or gum trees are of the genus *Eucalyptus* and are native to Australasia. The tallest trees in Australia are eucalyptus and the *E. regnans* is regarded as the tallest of the nonconifers anywhere in the world. Eucalyptus trees, notably the blue gum (*E. globulus*), were introduced into California in the 1880s, whereupon it immediately thrived. This same species was introduced into Ethiopia in the late 1800s, where it also has fared well and much accepted because of a general lack of timber in that area. Lines of eucalptus trees on the borders of citrus orchards in California are a common sight. It also has been reported that eucalyptus trees have been successful in reducing the occurrence of mosquitoes in marshy areas, the trees assisting in dewatering wet, boggy soil. There are some 600 species of eucalyptus, many of which are quite localized. Thus, generalization are difficult. Some of the more important species include:

Blue gum	*E. globulus*
E. caesia	
Cider gum	*E. gunnii*
Ghost gum	*E. papuana*
E. leucoxylon "Rosea"	
Lemon-scented gum	*E. citriodora*
E. miniata	
E. obliqua	
Red-flowering gum	*E. ficifolia*
E. regnans	
Red-flowering gum	*E. filifolia*
River red gum	*E. camaldulensis*
Snow gum	*E. niphophila*
Spinning gum	*E. perriniana*
Tasmanian snow gum	*E. coccifera*
Woodward's gum	*E. woodwardii*

The red-flowering gum is exceedingly showy when in bloom. Not a tall tree, this species seldom exceeds 50 feet (15 meters) in height. In season, masses of bright-red blossoms in large and heavy clusters nearly cover the tree. The clusters are from 6 to 10 inches (15.2 to 25.4 centimeters) across. Where trees are planted along pathways, the massive blossoms create clearing problems.

The blue gum has a blue-gray bark, quite smooth. The tree may exceed 300 feet (90 meters) in height. The leaf is dark green, 6 to 12 inches (15.2 to 30.4 centimeters) long and about $1\frac{1}{2}$ inches (3.8 centimeters) broad. Growth is rapid and long periods of drought can be withstood. It will survive temperatures as low as 20°F (4.4°C) for short periods. Although the best known of the species introduced into California, there are at least 80 species of eucalyptus growing in the state. Other species include: The white gum or manna gum (*E. viminalis*) with a gray-white bark and height of from 50 to 60 feet (15 to 18 meters). The leaf is from 4 to 8 inches (10 to 20.3 centimeters) in length. The tree grows fast and may have a girth of up to 3 feet (0.9 meters) within a 12-year period. The tree can tolerate poor soil. The dollar-leaf eucalyptus (*E. pulverulenta*) has blue-green leaves, essentially round in shape, and from 2 to 3 inches (5 to 7.6 centimeters) across. The tree grows to about 30 feet (9 meters) in height. The stalk is narrow, ranging up to about $2\frac{1}{2}$ inches (6.4 centimeters). The ornamental foliage is frequently used by florists and artists.

The Tasmanian snow gum is planted in England. The snow gum can withstand the climes of Mount Kosciusko (New South Wales) up to an altitude of 7000 feet (2100 meters).

The record eucalyptus (River red gum, *Eucalyptus camaldulensis*) growing in the United States is located in Kern County, California. As compiled by the American Forestry Association, this specimen has a circumference (at $4\frac{1}{2}$ feet; 1.4 meter above ground level) of 178 in (452 cm), a height of 171 feet (52.1 meters), and a spread of 68 feet (20.7 meters).

Eucalyptus oil, an essential oil, is derived from the leaves of several species of these trees. The oil is used in various medicinal and household preparations.

An Energy Tree? In 1902, the Secretary of Agriculture (Theodore Roosevelt Administration) commented, "The phenomenally rapid growth of the eucalyptus and the special adaptation of many species to dry climates render these trees of particular economic importance to the nation." In fact, the eucalyptus had many ardent supporters in the 1850s, shortly after its introduction into the United States. Unlike some trees and shrubs that have been introduced enthusiastically to the nation only to "backfire" by creating problems of their own (see **Casuarina; Kudzu; Melaleuca** for examples), the eucalyptus is still regarded in a positive way.

The latest recognized appeal of the eucalyptus is as a source of energy. Other industrial uses (as lumber and timber) in the United States have not been outstanding, but the Australians have been using the wood for decades in connection with shipbuilding, paper, fencing, and telegraph posts, among others. Among trees being studied in the United States as wood energy sources (because the day of the renewable energy resources will reawaken), numerous species of eucalyptus are being investigated. Objectives are to determine which species are able to grow quickly and dependably and in what geographical areas—so that transportation of wood fuel costs could be minimized. As observed by R. B. Pearce (*American Forests*, 30–34, January 1983), "Eucalyptus trees are ideal for short-rotation biomass production. The tree can establish a niche in any temperate climate, dry or wet, and is adaptable to the most improverished soil conditions. Most species can out-produce all other fuel stocks currently being studied (including many agricultural crops). The *Sydney bluegum* (*E. saligna*) can easily provide ten metric tons of oven-dried wood per acre (per 0.4 hectare) per year. And because most eucalyptus coppice (sprout from stumps), they can be harvested as often as twice a year without the need to replant. Unfortunately, a low tolerance to frost limits the distribution of the eucalyptus to California's coasts and lowland valleys, and to the warmer regions of Florida, Georgia, and the Carolinas."

EUCLASE. The mineral euclase is a silicate of beryllium and aluminum corresponding to the formula $BeA1SiO_4(OH)$, which crystallizes in the monoclinic system. It has a perfect prismatic cleavage; hardness, 7.5; specific gravity, 3.1; luster, vitreous; is colorless to sea-green or blue. It has been used to a very slight extent for jewelry as its transparent crystals somewhat resemble the aquamarine. Euclase occurs in the Minas Gerais region, Brazil, associated with topaz and beryl, and also in the Ural Mountains, where it is found in gold-bearing sands. The name euclase is derived from the Greek, meaning easiness and fracture, in reference to its easily cleaved crystals.

EUCLIDEAN GEOMETRY. See **Geometry.**

EUCLIDEAN SPACE. A generalization of the algebraic, geometrical and topological properties of the line, the plane, and three-dimensional space to n dimensions.

A Euclidean n-space is an n-dimensional linear space provided with a scalar product. A concrete realization of Euclidean n-space is the set of n-tuples $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ of real numbers with the scalar product of $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and $(\mu_1, \mu_2, \ldots, \mu_n)$ defined as $\lambda_1\mu_1 + \lambda_2\mu_2 + \ldots + \lambda_n\mu_n$.

In mathematical models for physical systems, it is often necessary to consider spaces having more than three dimensions; for example, six dimensions are needed to locate two particles in Euclidean three space.

EUCLID'S ALGORITHM. See **Algorithm.**

EUDIOMETER. A graduated tube closed at one end in one form of which two platinum wires are sealed so that a spark may be passed through the contents of the tube; used to measure the volume changes in the combustion of gases.

EULER ANGLE. One of three parameters describing the orientation of a rigid body relative to a Cartesian coordinate system (x, y, z)

fixed in space. Suppose another coordinate system (x', y', z') is fixed in the body. Then the two systems may be made coincident by three successive rotations, applied in the appropriate order, and the three angles are the Euler angles.

See also **Coordinate System.**

EULERIAN COORDINATES. Any system of coordinates in which properties of a fluid are assigned to points in space at each given time, without attempt to identify individual fluid parcels from one time to the next. Since most observations in meteorology are made locally at specified time intervals, an Eulerian system is usually, though by no means always, more convenient. A sequence of synoptic charts is an Eulerian representation of the data.

EULER-MASCHERONI CONSTANT. A number, also often called simply Euler's constant, which occurs in one definition of the gamma function. It can be defined by several equivalent infinite integrals, one example being

$$C = \int_{\infty}^{0} e^{-t} \ln t \, dt$$

Its numerical value is 0.577215665 . . . but the quantity $\gamma = 1.781072$ . . . , defined by $\ln \gamma = C$ is sometimes defined as the Euler-Mascheroni constant.

See also **Gamma Function.**

EUPHORBIACEAE. Also known as the spurge family, this is a large family of usually cactus-like trees and shrubs which often yield a milky juice or latex. They occur in many tropical and temperate regions. There are a few herbaceous species, especially in cooler regions. Many of the tropical species are interesting xerophytes, plants capable of enduring the driest climates. Often these have a habit very similar to that of species of cactus, with which they may easily be confused, especially if not in flower. However, nearly all members of the spurge family contain a milky juice which exudes from them when the surface is cut or broken. This milky juice, or latex, will readily distinguish them from cacti, which lack latex. Leaves, when present, are usually alternate and have stipules. In many species, such as the frequently cultivated *Euphorbia splendens*, or "crown of thorns," the leaves soon drop off, leaving a spine-covered stem. In one genus, *Phyllanthus*, leaves are frequently reduced to minute scales, and the stem flattened and green; in these the small pinkish flowers are borne around the edge of the flattened stem. In some species of *Euphorbia* the leaves near the top of the stem become brilliantly colored, as in *Euphorbia pulcherrima*, the poinsettia, where they are bright red. Such leaves, surrounding the inconspicuous flower masses, are often mistaken for parts of the flower.

The inflorescence, in members of this family, is often very complex. In many species the individual flowers are crowded together in such a way as collectively to resemble a single large flower. The flowers are unisexual. The plants are either monoecious or dioecious. In many cases the flowers entirely lack both calyx and corolla, in others a calyx is present, but no corolla, while in some both calyx and corolla are present. They are regular flowers with the perianth commonly 5-parted. The number of stamens varies from one to many; in many cases they are variously united; in some, as in the castor bean, they are branched. The ovary is usually 3-celled, with one or two ovules in each cell. The fruit is a capsule, which when mature often opens with considerable force, throwing the seeds out, often to considerable distances. The seeds have an abundant endosperm and a caruncle.

Among the members of this family are some plants of great economic importance. Many others are poisonous plants.

Hevea brasiliensis, the Para rubber plant, is perhaps the most valuable member. Species of *Ricinus* supply castor oil. Species of *Manihot*, a South American genus, yield cassava or mandioc, a starchy foodstuff, prepared from the large roots. *Manihot esculenta*, for instance, has long been cultivated in Brazil. It is a large, somewhat bushy herb with long-petioled leaves, the smooth blades of which are deeply cleft into 3–7 lobes. The roots, which have the appearance of sweet potatoes, are eaten in much the same way as sweet potatoes. Grated, they yield a starchy product used like bread. From the roots tapioca may be prepared. The poisonous principle which is present in many species of *Manihot* is removed by squeezing or destroyed by heating. From other species of this genus may be obtained, by tapping, a milky juice which is a source of rubber. See also **Rubber (Natural).**

Hura crepitans, the sand box tree, is another member of the family of some slight commercial value. The plant is a fairly large tree the stem of which is covered with short, sharp spines, and bears long-petioled toothed leaves. The fruit is composed of numerous hard carpels which, when mature, explode violently, throwing the seeds out forcibly. These fruits, about 3 inches (7.5 centimeters) in diameter, were formerly gathered and wired to prevent bursting. When dry they were used as containers for the fine sand which was then used to blot ink—hence the common name of the tree. The wood of the tree is used locally, but rarely exported. The milky juice of the tree is very poisonous. This juice, mixed with meal or similar substances, and thrown into the waters of a stream or lake, stupefies the fish present therein, so that they may be readily captured. The poison does not render the fish unfit for human consumption.

Several species of *Croton* yield important purgative drugs. *Croton eluteria* gives Cascarilla bark, used as a tonic.

Jatropha curcas, a small shrub or tree, bears egg-shaped green fruits which contain a high percentage of an odorless oil called "curcas" oil, used in making paints, as a lubricant, and in soapmaking. From the leaves of this tree, natives of the Philippine Islands prepare a fish poison used in much the same way as that obtained from *Hura*.

Aleurites are tropical trees bearing small many-seeded fruits extremely rich in oil. *Aleurites triloba*, the candlenut of the orient, produces a fruit extensively used for food and for light. *Aleurites cordata*, a native of China, is the "varnish-tree." *Aleurites fordii* yields tung oil.

EUPHOTIC ZONE. The layer of a body of water which receives ample sunlight for the photosynthetic processes of plants. The depth of this layer varies with the water's extinction coefficient, the angle of incidence of the sunlight, length of a day, and cloudiness; but it is usually 80 meters or more. The depth of compensation is the lower boundary of the euphotic zone. See also **Ecology.**

EUROPEAN CORN BORER (*Insecta, Lepidoptera*). A small moth, *Pyrausta nubilalis*, whose larva bores in the stems of plants, especially Indian corn. The species is closely related to certain North American moths and was first noticed as a pest in the eastern part of the United States about 1920. Since then it has spread westward.

EUROPIUM. Chemical element symbol Eu, at. no. 63, at. wt. 151.96, sixth in the Lanthanide Series in the periodic table, mp 822°C, bp 1529°C, density 5.245 g/cm^3 (20°C). Elemental europium has a body-centered cubic crystal structure at 25°C. The pure metallic europium is silver-gray in color under vacuum, but oxidizes readily in air and must be handled in an inert atmosphere. Europium is very soft as compared with the other rare-earth metals. Two stable isotopes of the element occur naturally ^{151}Eu and ^{153}Eu. Upon absorption of thermal neutrons, ^{151}Eu forms ^{152}Eu with a half-life of 13 years; ^{153}Eu forms ^{154}Eu with a half-life of 16 years. The latter further decays to ^{155}Eu with a half-life of 1.7 years. In terms of abundance, europium is present on the average of 1.2 ppm in the earth's crust, making its potential availability greater than antimony, bismuth, or cadmium. The element was first identified by Sir William Crookes in 1889. Europium dissolves readily in dilute mineral acids and reacts with H_2O at room temperature. The metal is not known to be toxic but because of its high reactivity in air, great care in handling is mandatory. Electronic configuration

$$1s^2 2s^2 2p^6 3s^2 2p^6 3d^{10} 4s^2 4p^6 4d^{10} 4f^6 5fs^2 5p^6 5d^1 6s^1$$

Ionic radius Eu^{2+} 1.09 Å, Eu^{3+} 0.950 Å. Metallic radius 1.995 Å. First ionization potential 5.67 eV; second 11.25 eV. Oxidation potential $Eu^{2+} \rightarrow Eu^{3+} + e^-$, 0.43 V. Other important physical properties of europium are given under **Rare-Earth Elements and Metals.**

Europium occurs in the rare-earth fluocarbonate mineral bastanite, mainly found in southern California. The mineral contains between 0.09 and 0.11% Eu_2O_3. Other minerals, such as xenotime and monazite, also contain europium compounds and sometimes are used as sources of the element.

Because of the desirable nuclear properties of the element, europium has received serious consideration for the construction of nuclear reactor hardware. Earlier commercial unavailability of the element, however, favored the use of other materials. Some small reactors have been constructed in which europium molybdate has been the major control-rod component. With much increased availability of the metal in recent years, the prospect of further usage of europium in reactor design are good. A europium-activated yttrium orthovanadate, Eu:YVO_4, has shown promise as a red phosphor for commercial television. An increase of 40% in light output has been claimed. With this system, the average color television set would require about $\frac{1}{2}$ g of Eu_2O_3 and 6 g of Y_2O_3. The stimulus resulting from this discovery resulted in the development of other new phosphors involving europium in various host matrices. These new materials have been used in high-intensity mercury-vapor lamps, general-purpose fluorescent lamps, x-ray screens, charged-particle detectors, and neutron scintillators. In some optically-read memory systems, ferromagnetic europium chalcogenides (sulfides, selenides, and tellurides) have been used. Other electronic and semiconductor uses of europium are under serious investigation.

NOTE: This entry was revised and updated by K. A. Gschneidner, Jr., Director, and B. Evans, Assistant Chemist, Rare-Earth Information Center, Energy and Mineral Resources Research Institute, Iowa State University, Ames, Iowa.

EUSTACHIAN TUBE. A slender canal between the pharynx and the middle ear of vertebrates. It permits the equalization of pressure on the two surfaces of the eardrum. See also **Hearing and the Ear.**

EUTAXIC. A term proposed by Keyes in 1901 for obviously stratified sedimentary ore deposits as contrasted with those which are unstratified. The latter he designated as ataxic.

EUTECTIC. An eutectic reaction is a reversible isothermal transformation in which, during cooling, a single liquid phase is transformed into two or more solid phases, the number of solid phases being equal to the number of components. In a given alloy system, at a fixed pressure, all phases will have fixed compositions during the isothermal transformation. The temperatures at which the freezing occurs is known as the eutectic temperature, while the composition of the liquid phase is called the eutectic composition. On a temperature-composition binary phase diagram, the eutectic point is determined by the eutectic composition and the eutectic temperature. In general, an alloy of the eutectic composition freezes at a minimum temperature. For this reason, eutectic compositions, or compositions close to the eutectic, are frequently used in low melting point solders.

By a similar usage in petrology, a eutectic is a discrete mixture of two or more minerals, in definite proportions, which have simultaneously crystallized from the mutual solution of their constituents. The eutectic point is the lowest temperature at any given pressure at which the above physical-chemical process may take place. The eutectic ratio is the ratio by weight of two minerals which originate by the above process.

EUTECTOID. This is a phase transformation analogous to an eutectic where a single solid phase, instead of a liquid phase, is transformed into two or more different solid phases. The number of solid phases in the resulting eutectoid structure is equal to the number of components in the system. Under very slow cooling, the eutectoid transformation should occur at the eutectoid temperature. However, due to the sluggishness of solid state transformations, there is usually some hysteresis with the transformation temperature depressed on cooling and raised on heating. Under equilibrium conditions, the compositions of the various phases are fixed in an eutectoid reaction just as they are in an eutectic transformation.

The best known eutectoid reaction is that which occurs in steel where the austenite phase, stable at high temperatures, transforms into the eutectoid structure known as pearlite. In this transformation, the austenite phase, containing 0.8% carbon in solid solution, transforms to a mixture of ferrite (nearly pure body-centered cubic iron) and iron-carbide (Fe_3C). At atmospheric pressure, the equilibrium temperature for this reaction is 723°C. This temperature is the eutectoid temperature.

In binary alloy systems, a eutectoid alloy is a mechanical mixture of two phases which form simultaneously from a solid solution when it cools through the eutectoid temperature. Alloys leaner or richer in one of the metals undergo transformation from the solid solution phase over a range of temperatures beginning above and ending at the eutectoid temperature. The structure of such alloys will consist of primary particles of one of the stable phases in addition to the eutectoid, for example ferrite and pearlite in low-carbon steel. See also **Iron Metals, Alloys, and Steels.**

EVAPORATION. The evaporation of a liquid consists in the escape from the main body of the liquid of those molecules which, in their thermal agitation, are moving with a sufficient speed to break through the surface tension; that is, whose kinetic energy exceeds the work function of cohesion at the surface. Since only a small proportion of the molecules are at any instant located near enough to the surface and are moving in the proper direction to escape, the rate of the evaporation is limited. It is easy to see why it proceeds more rapidly with higher temperature, and why liquids of low surface tension are relatively volatile. Also, as the faster moving molecules emerge, those left behind have less average energy, and the temperature of the liquid is thereby lowered. If the evaporation takes place in a closed vessel, the escaping molecules accumulate as a vapor above the liquid. Many of them return to the liquid, such returns being more frequent, the greater the density and pressure of the vapor. Presently the processes of escape and return come to equilibrium; the vapor is then said to be "saturated," its density and pressure no longer increase, and the cooling effect ceases. Even a warm breeze cools the skin because it removes the evaporating perspiration and prevents saturation.

Evaporation is a major chemical engineering unit operation for bringing about separations of liquids and solids and, in particular, to recover the solute (such as a dissolved salt) from the solvent (frequently water). Usually, the main object of the separation is the solute. The pulp and paper industry is a large user of evaporation equipment. In pulp mills, after the digestion system, the pulp is leached with water and the chemical solids are dissolved out almost completely by a pulp-washing system. The recovered liquid from these operations is fed to an evaporator, generally at about 15% total dissolved solids content. The evapo-

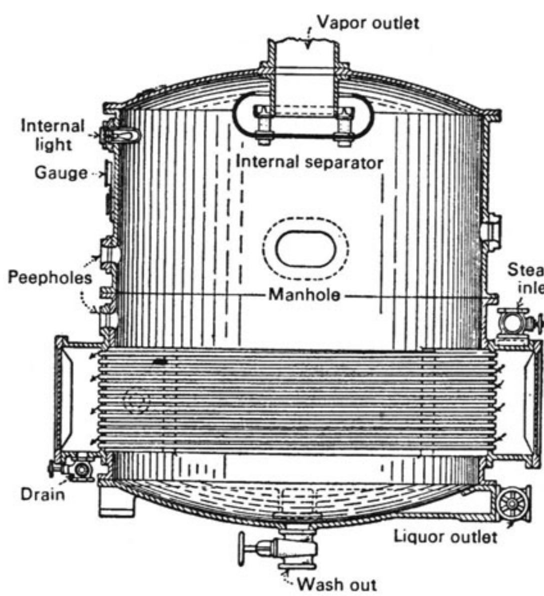


Fig. 1. Horizontal-tube evaporator.

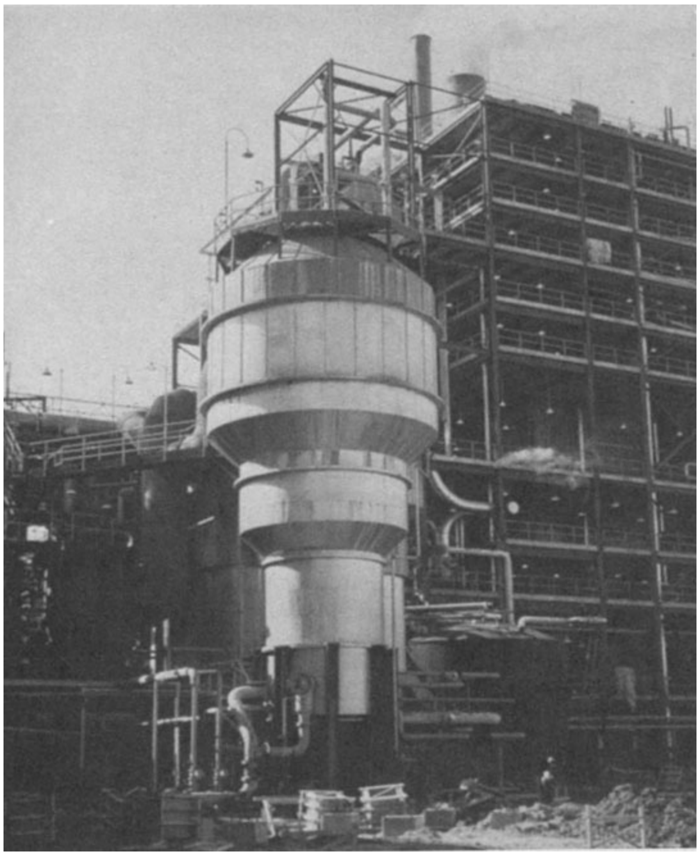


Fig. 2. The sixth-effect evaporator body of a multiple-effect evaporation system.

rator removes much of the water and in so doing concentrates the liquid to 55–65% total dissolved solids, whereupon the solution then can be further processed in a chemical recovery furnace. Other types of pulp processing also involve chemical-containing solutions which must be evaporated for recovery of valuable chemicals. Evaporation also is used extensively in the production of table and industrial salt (sodium chloride) as well as other salts, in caustic-chlorine production, in the phosphate industry, and in food processing. Evaporators can be large structures as the illustrations indicate.

Evaporation is in principle the same operation as plain distillation, with the modifications in practice that (1) the vapor may or may not be recovered, (2) the residue in the evaporator may or may not contain solids, and (3) vacuum evaporation is frequently used in a single compartment or in multiple stages with each successive stage operated at an increasing vacuum utilizing the heat of condensation of the vapor from the preceding stage. In multiple stage evaporators there is a saving in the cost of heat and an increased expenditure for apparatus. Vacuum evaporation is frequently utilized to lower the temperature to which a substance is subjected and thus avoid decomposition by passing a current of warm dry air over the substance. Combined high-vacuum and very low temperature evaporation or drying is practiced in the final removal of water vapor from frozen penicillin, due to the heat-sensitive nature of this material. Water vapor passes from the place of higher concentration, that is, the substance, to the place of lower concentration, that is, the air, and is thus removed from the substance. If oxygen of the air reacts with the substance, an inert gas such as nitrogen may be substituted for air.

An evaporator system may be single effect (Fig. 1), in which the steam is produced from one evaporator, or multiple effect, in which the steam is produced from several evaporators in series. In a multiple effect system the vapor from one evaporator becomes the heating steam in the succeeding. Unusual conditions met in industrial or steam heating plants may require so large a fraction of make-up as to warrant double, triple, or quadruple effect evaporators. The central generating station ordinarily employs single effect and rarely requires more than a double effect system. The ratio vapor produced/steam used is about 0.8 for the single effect, 1.5 for the double effect, and 2.5 for the triple effect system. See Figs. 2 and 3. Evaporator feed is sometimes preheated to increase evaporator capacity.

Evaporators are classed as film, flash, or submerged-tube types. The first and last are steam-tube types; in the former the raw water trickles over the hot tubes, in the latter the tubes are entirely surrounded by the water being evaporated. The flash type produces steam by dropping the pressure on water at the saturation temperature. The excess heat flashes part of the water into steam, then the remainder is drawn off, reheated, and again flashed.

EVAPORITE. A sedimentary rock formed by precipitation from waters at the earth's surface. As described by Lowenstein (*Science*, 1090, September 8, 1989), ancient evaporites have been used to track the chemistry of ancient surface waters, particularly seawater. Study of marine evaporites has led to the general (not unanimous) conclusion that

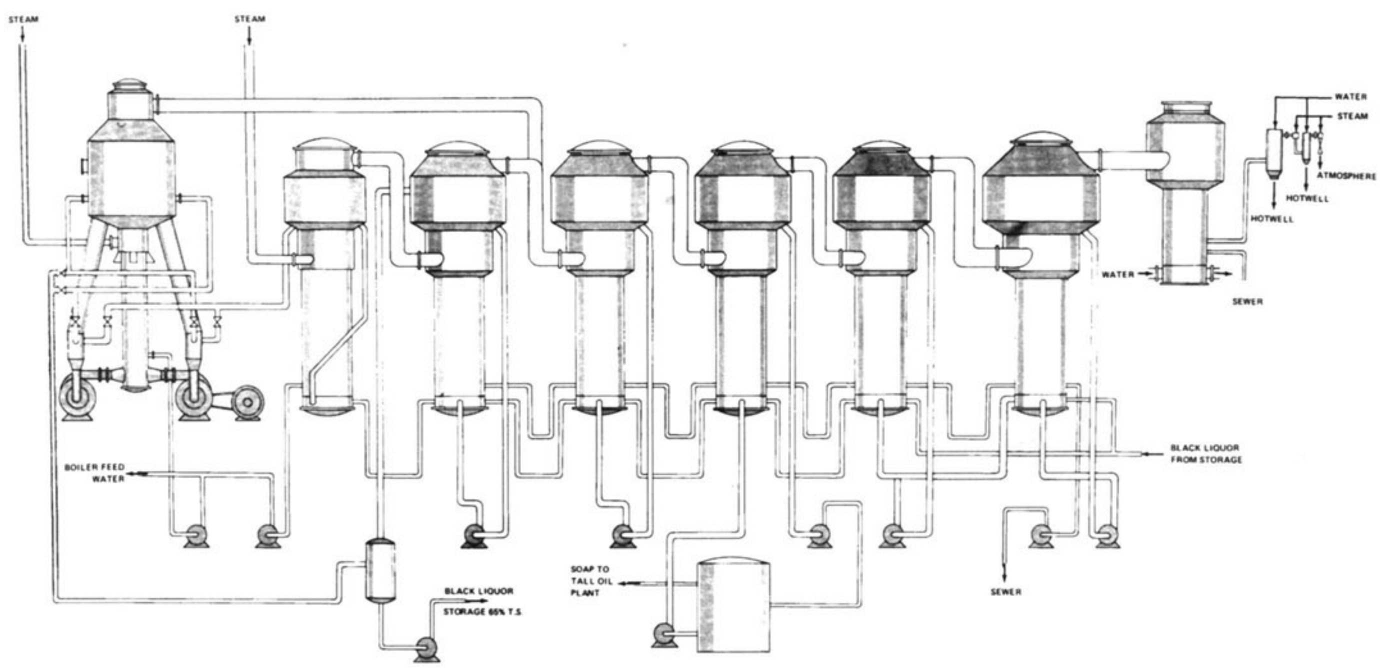


Fig. 3. Compound multiple-effect kraft mill evaporator.

the major elemental chemistry of seawater has not changed significantly during the last 600 million years.

EVECTION. A perturbation of the moon in its orbit due to the attraction of the sun. This results in an increase in the eccentricity of the moon's orbit when the sun passes the moon's line of apsides and a decrease when perpendicular to it. Evection amounts to 1 degree 15 minutes in the moon's longitude at maximum.

EVEN-EVEN NUCLEI. Atomic nuclei that contain an even number of protons and an even number of neutrons.

EVENT. A happening, represented by a point (x, y, z, t) in the spacetime continuum. It is a fundamental assumption of the theory of relativity that all physical measurements reduce to observations of relations between events.

EVOLUTE. The evolute of a given curve is the locus of its centers of curvature. See also **Involute.**

EVOLUTION (Biological). (1) The development of an organism toward perfect or complete adaptation to environmental conditions to which it has been exposed with the passage of time; (2) the theory that life on earth developed gradually from one or several simple organisms (appropriate molecules) to more complex organisms. Sometimes called *organic evolution*. The term *evolutionary biology* is also used.

In contemplating the great diversity and vast numbers of species of life on earth, during post-Renaissance times, Linnaeus (1707-1778) and many of his contemporaries ascribed these to a concept of special creation—"there are just so many species as in the beginning the Infinite Being created."

In his *Philosophie Zoologique*, the French philosopher, Jean Baptiste de Lamarck, observed in 1809, "the existence in organisms of a built-in drive toward perfection; the capacity of organisms to become adapted to 'circumstances' [environment in modern terminology]; the frequent occurrence of spontaneous generation; and the inheritance of acquired characters, or traits."

The fourth of the foregoing observations was shown to be inaccurate, but about a half-century later, Darwin (1859) accepted the concept, i.e., "assumed that the use or disuse of a structure by one generation would be reflected in the next generation." Some other theorists who followed Lamarck and Darwin also accepted the validity of Lamarck's fourth observation and it remained for the German biologist, August Weissman (1834–1914), to stress the impossibility (or at least the improbability) of that observation. This constituted the first of several modifications of early hypotheses and the later theory, progressing from Lamarckism to Darwinism, to neo-Darwinism, and to the more recent synthetic theory and to the current conceptual developments.

Seldom stressed is the fact that although the theory of evolution is some 120 years old (Darwinism) or 170 years old (Lamarckism), the theory, in terms of research required to round out its full development and implications, is still in its infancy. Compared with the life sciences information bank of the early 1980s, the early theoretical endeavors were conducted in a scientific vacuum. Thus, the emergence of advanced genetics, cytology, molecular biology and numerous other sciences relative to evolutionary biology have impacted and will continue to impact on those scholars who pursue the theory in their efforts to construct a continuum of events that led from a lifeless earth to living organisms and systems as we know them today.

The difficult tasks facing investigators in this field today are typified, in part, by the observations of one authority on the concept of chemical evolution of life: "The evolution of the genetic machinery is the step for which there are no laboratory models; hence one can speculate endlessly, unfettered by inconvenient facts. The complex genetic apparatus in present-day organisms is so universal that one has few clues as to what the apparatus may have looked like in its most primitive form." (R. E. Dickerson, 1978)

During 1980s, there was a trend toward more effective comingling of various scientific disciplines in an effort to weave a tighter and more coherent network of information concerning the theory of evolution. These include, among many others, the improved integration of findings of paleontologists, archeologists, anthropologists—and chemists and physicists who have developed improved age-estimating and dating techniques. As pointed out by Woodruff (1980) in a review of Stanley's book (see reference list): "Paleontology is currently undergoing an exciting rejuvenation, and Stanley and his fellow paleobiologists (as they are now called) have introduced some scientific rigor into a traditionally descriptive field. Now, in place of inspired speculation, we see attempts to test hypotheses derived from theoretical population ecology against the extensive fossil record. . . . Evolutionary biologists can no longer ignore the fossil record on the ground that it is imperfect."

Possibly the most important conference on evolutionary biology held since the 1940s convened in Chicago in the fall of 1980. As reported (Lewin, 1980), the principal issues discussed at the meeting were (1) the tempo of evolution; (2) the mode of evolutionary change; and (3) the constraints on the physical form of new organisms. Dominating the field of evolutionary biology for several decades, the Modern Synthesis concept of evolution (so named by Julian Huxley in 1942) was reexamined in the light of many intervening years of progress in the biological sciences. In essence, this concept assumes that the pace of evolutionary change is slow, that the direction of evolutionary change is governed by natural selection involving small variations, and that the variants that survive are those that are environmentally superior. The proceedings are well encapsulated by Lewin. As one scientist at the conference observed, "I hope that this meeting will lead to a reapproachement. I hope it will set the basis for a reconstruction of ideas." A number of interesting developments have since occurred.

Punctuated Equilibrium. In 1972, Eldredge (American Museum of Natural History) and S. J. Gould (Harvard University) proposed the hypothesis of punctuated equilibrium. In explaining gaps in the fossil record, the hypothesis proclaims that the patterm of stasis and abrupt changes apparent in the record is real and not an artifact of its incompleteness. This is based on the assumption that once a species has arisen, it remains essentially unchanged for most of its history, but when changes do occur, they do so swiftly. Also, it contrasts with other hypotheses to the effect that changes for the most part resulted from a steady accumulation of small modifications. The pattern of change accordingly would be gradual. The fossil record does not show this. The two views obviously are controversial and, as of the late 1980s, were not resolved. As reported by Lewin (1986), several scholars have created mathematical models, which involve equations that describe the dynamics of diffusion. These models suggest that the pattern of punctuated equilibrium—stasis and rapid change—is predictable from Neo-Darwinian theory. The context of the pattern is the absence of environmental change. The models indicate that the pattern is evolution by "jerks," as it has been termed, during periods of environmental constancy.

Function and Adaptation. In evolutionary biology studies, it is frequently difficult to separate function from adaptation. Concerning neuronal circuits, for example, Dumont and Roberston (1986) observe that it may not be possible to explain many features of nervous systems in terms of adaptive significance. Rather, it may be more appropriate to consider how a neural circuit, or any other feature, is shaped during evolution. The effects of evolution can be considered to be influenced by four types of determinants: (1) adaptive influences, which are directly related to optimization of the effect of the behavior; (2) developmental constraints, which pose restrictions on the final form of the nervous system; (3) historical influences, by which the form of the present-day nervous system reflects the ancestral form; and (4) certain architectural features, which are imposed by the materials and design of the organism. Although this classification of determinants is by no means perfect, adaptation clearly does not act alone to shape a circuit during evolution. Recently, evolutionary biologists have argued for the importance of such nonadaptive processes in evolution. Gould (1982) pointed out that the brain, because of its complexity, is perhaps the most striking example of the effects of nonadaptive processes in evolution.

Morphology and Molecular Biology. The interdisciplinary nature of modern evolutionary biology was apparent at the July 1985 meeting of the Third International Congress of Systematic and Evolutionary Biology held in Brighton, England. As reported by Lewin (1985), there was a symposium on Molecules versus Morphology, which interacted the molecular biologist with the morphologist. Phylogenetic reconstruction requires the search for signs of shared ancestry, specifically the identification of homologous characters that uniquely link two or more species as an evolutionarily derived group. This search, traditionally, has been for morphological structures. It now includes molecular sequences of proteins and DNA. Molecular and morphology data differ significantly as regards rate of change. If anatomical structures alter during their adaptation at least partly in response to pressures of natural selection, pressures which may differ markedly in time and magnitude, then no theoretical argument can be made for regularity of change. From the viewpoint of morphology, modification can take place in spurts, or slowly in gradual responses. In contrast, speculation and some empirical evidence indicates that molecular change may occur with considerable constancy, as guided by a so-called molecular clock. Because of this dichotomy, much additional research and indeed cooperation between the molecular biologists and morphologists will be required to bring unity out of these differing views and experiences to date.

Evolution in the Broad Sense. In a practical sense, disregarding the challenge and fascination of the theory, the primary usefulness to date of the theory of evolution has been its role as a unifying concept for the biological sciences as we know them today—not dissimilar to the role played by the earlier (Stahl, 1600—1734) phlogiston theory (later supplanted by the laws of thermodynamics) which brought a degree of order to physics.

In essence, the various theories of cosmogony (origin of galaxies, stars, planets, etc.) furnish the prelude to the theory of evolution. In commenting on the relative chores of the cosmogonist and the evolutionary biologist, Mayr (1978) observed: "For one thing, it (biological evolution) is more complicated than cosmic evolution, and the living systems that are its products are far more complex than any nonliving system." Over the future years of continuing investigation and conceptualization, the efforts of these two fields nevertheless require coordination and tight information transfer because the theory of evolution is time sensitive even if in terms of billions of years—because the best estimated age of the earth, of which only the last portion has embraced environments suitable to nurture and sustain life, must encompass all of the events described by the evolutionary biologist.

The many activities which investigators have undertaken in the past and are undertaking today to further mold and refine the theory of evolution are too numerous, complex, and detailed to report here. Several references are listed at the end of this entry for further reading.

Related topics include **Cell (Biology)**; and **Fossils and Paleontology**.

Additional Reading

Cech, T. R.: "RNA as an Enzyme," *Sci. Amer.*, 64 (November 1986).
Garrett, W. E.: "Where Did We Come From?" *Nat'l. Geographic*, 434 (October 1988).
Rennie, J.: "In the Beginning," *Sci. Amer.*, 28 (September 1989).
Rubenstein, E.: "Stages of Evolution and Their Messengers," *Sci. Amer.*, 132 (June 1989).
Simons, E. L.: "Human Origins," *Science*, 1343 (September 22, 1989).
Stringer, C. B.: "The Emergence of Modern Humans," *Sci. Amer.*, 98 December 1990).
Waldrop, M. M.: "Did Life Really Start Out in an RNA World?" *Science*, 1248 (December 8, 1989).
Waldrop, M. M.: "Spontaneous Order, Evolution, and Life," *Science*, 1543 (March 30, 1990).
Waldrop, M. M.: "The Golden Crystal of Life," *Science*, 1080 (November 23, 1990.

EVOLUTION (Mathematics). See **Square and Square Root.**

EXCAVATION. See **Earthwork.**

EXACERBATION. When used in a medical description, exacerbation means an intensification (possibly arising from irritation, aggravation of underlying causes) of pain and other symptoms of a disease or disorder that, in chronic situations, may occur intermittently and rise above a general background of tenderness, discomfort, etc.

EXCHANGE DEGENERACY. An exchange process which does not entail a change in value or configuration. For example, by the Heitler-London theory, the essential reason for the strong attraction (or repulsion), of the two H-atoms in the H_2 molecule is the exchange degeneracy, i.e., the fact that for very large internuclear distance, by exchange of the two electrons of the two atoms a configuration results that is indistinguishable from the original configuration. Therefore, as they approach, there arises an interaction between them which may be treated mathematically as electron exchange.

EXCHANGE ENERGY. A specifically quantum-mechanical effect which has no classical analog. It is due to the interaction between two systems that arises, or could arise from the continuous exchange of a particle between them.

Suppose, for example, that two electrons are in states that allow them to come close together. Then, because they are indistinguishable particles, one could not tell the difference if they exchanged states. That is, one must combine with the original description (i.e., wave-function) a function in which the electrons have actually changed places, it can easily be shown that two such combined states are possible—the symmetric and antisymmetric combinations—and that in each of these the energy is significantly different from that of the original state. Exchange energy is the origin of covalent bonding, of ferromagnetism and antiferromagnetism, probably of nuclear forces (where exchange energy could arise by exchange of π-mesons between nucleons, giving rise to an effective potential which involves an operator which exchanges the spins, isotopic spins and/or positions of the particles) and of numerous other physical phenomena.

EXCHANGE FORCES. Nonclassical, quantum mechanical forces that arise from the phenomena of exchange and that account for the exchange energy. The binding of a hydrogen molecule and the covalent bonding in molecules can be looked upon loosely as due to the exchange of electrons among the atoms. The Coulomb force between charged particles can be looked upon (in quantum electrodynamics) as due to the exchange of photons. The nuclear force between nucleons can be looked upon as due to the exchange of charged or neutral π-mesons.

EXCHANGE (Particle). 1. A quantum mechanical concept based on the idea of identical particles, which are particles having the same intrinsic properties, such as rest mass, spin and charge. For example, suppose that two electrons are in states that allow them to come close together. Then, because they are indistinguishable particles, one could not tell the difference if they exchanged states. Thus the wave function of the system must be such that an exchange of the electrons leaves the magnitude of the wave function unchanged, except possibly for sign, i.e., the wave function must be either symmetric or antisymmetric to an exchange of the two particles. Particles whose total wave function (including both space and spin coordinates) is symmetric under an exchange operator obey the Bose-Einstein statistics. Particles whose total wave function is antisymmetric obey the Fermi-Dirac statistics.

2. Exchange is also used more specifically as the exchange of one particle between two others, as in the exchange of the single electron between the two identical protons in the hydrogen molecular ion, or the exchange of a meson between two nucleons.

Some concepts have been altered during recent years. Check entry on **Particles (Subatomic).**

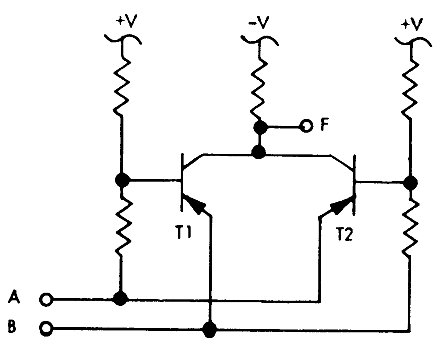
EXCITATION. This term has three common uses in physics and engineering: 1. Addition of energy to a system, whereby it is transferred from its ground state to a state of higher energy, called an excited state. 2. The field excitation of dynamo machines, meaning the current or voltage of the field circuit. 3. In vacuum-tube and transistor circuits, the input signal of any stage is commonly called the excitation. Thus in a radio receiver, the signal picked up by the antenna supplies the excitation for the first state, the output of the first supplies the excitation for the next, and so on.

EXCITATION CURVE. In nuclear physics, a graphical relationship between the energy of the incident particles or photons, and the relative yield of a specified nuclear reaction.

EXCITER. This term has four common uses in engineering: 1. In antenna terminology, the portion of a transmitting array of the type which includes a reflector, which is directly connected with the source of power. 2. In transmitters, the oscillator which supplies the carrier or subcarrier frequency voltage to drive the stages which ultimately lead to the final power output stage. In FM systems this unit includes all the frequency generating, modulating and frequency multiplying circuits of the transmitter. 3. In photoelectric reproduction of film, the lamp which supplies a light source of constant amplitude. 4. A generator used to supply the field currents of a larger direct current generator or of an alternator.

EXCITING CURRENT (Transformer). This is the current which supplies the core losses and magnetizing current for a transformer. When the transformer is supplying a load this excitation current is one component of the total input, the other being the component which balances the load current. The current establishes the magnetic field in the core. furnishing energy for the no-load power losses in the core. Also called *magnetizing current*.

EXCLUSIVE OR CIRCUIT. A logical element which has the properties that if either of the inputs is a binary 1, then the output is a binary 1. If both the inputs are a binary 1 or 0, the output is a binary 0. In terms of Boolean algebra, this function is represented as $F = AB' + BA'$, where the prime denotes the **NOT** function. With reference to the transistor exclusive **OR** circuit shown in the accompanying diagram, the output is positive when either transistor is in saturation. When input A is positive and B is negative, transistor T_2 is in saturation. When B is positive and A is negative, transistor T_1 is in saturation. When A and B are either both positive or both negative, then both transistors are cut off and the output F is negative.



Exclusive OR circuit.

Although shown as discrete devices in the diagram, fabrication using large-scale integrated circuit technology may utilize other circuit and device configurations.

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

EXCRETION. The removal of the waste products resulting from the chemical transformation of materials in the body.

The oxidation of materials derived from foods for the release of energy may produce carbon dioxide and water whether the compound oxidized is a protein, a carbohydrate, or a fat, but since proteins contain nitrogen and other elements in addition to carbon, hydrogen, and oxygen, they also give rise to more complex waste products. The chief nitrogenous wastes of animals are urea and uric acid. The elimination of all of these compounds and other substances of like derivation is excretion.

Many small animals, including both protozoans (*Protozoa*) and more complex forms, apparently discharge these wastes from the surface of the body generally, while in others a special excretory system occurs. Even in those forms which have a complex excretory system, any moist surface directly or indirectly exposed to the medium surrounding the animal is favorable for the diffusion of materials into or out of the body, and so may carry on excretion. The wastes passed out in this manner are largely carbon dioxide and water, although the discharge of water may also take out dissolved solids. Thus the lungs of a terrestrial vertebrate eliminate carbon dioxide and the sweat glands of the skin of some animals discharge water with other materials, including nitrogenous wastes, in solution. By far the greater part of the complex wastes is eliminated by the excretory system.

In complex animals other organs than those directly involved in the elimination of wastes may play an important intermediary role. The circulatory system of the vertebrate, for example, transports all wastes from the tissues where they are formed to organs which act upon them and finally to the centers which remove them from the body. The liver removes some substances, including complex organic compounds resulting from the destruction of old red blood cells, and discharges them in the bile by way of the intestine. It also transforms ammonia and amino acids, resulting from the oxidation of proteins, into urea which is returned to the blood to be removed by the kidneys.

See also **Urine.**

EXCRETORY SYSTEM. An organic system whose principal or only function is the removal of complex wastes from the animal body. The excretory system of humans is discussed under **Kidney and Urinary Tract.**

Some animals lack an organized excretory system, discharging wastes from the surface of the body generally, but others, even among the one-celled animals, have special excretory structures. The contractile vacuoles of protozoans are supposed to carry out this function.

In the flatworms a special excretory system based on the flame cell appears. Flame cells are large and hollow, with a group of cilia projecting into the cavity whose movement drives out the liquid discharged by the cell. The cavity of each flame cell joins a small duct and these ducts converge to form larger ducts which ultimately open at the surface of the body. Flame cells emptying by ducts into a vesicle connected with the caudal end of the alimentary tract also occur in rotifers.

Roundworms have two slender excretory canals along the sides of the body which unite to empty by a single pore near the anterior end.

In the segmented worms, the body cavity becomes involved in excretion. Two forms of tubes, the coelomoducts and nephridia, open from the coelom to the exterior in these worms. These organs are segmentally arranged ciliated tubes with a funnel-shaped inner end and a minute opening externally. They are variously associated in the excretory organs of different species and in some do not open into the coelom but are provided with cells much like flame cells which are called solenocytes.

Arthropods of different classes have special excretory structures, including the coxal glands of scorpions, said to be derived from coelomoducts, and the Malpighian tubules of insects. The latter are slender tubules opening into the alimentary tract at the caudal end of the stomach and blind at their other end.

The occurrence of solenocytes in the lancelets of the phylum *Chordata* is unusual in this phylum, since in the true vertebrates a pair of kidneys are the chief excretory structures. They are developed from intermediate mesoderm. The excretory unit in these organs is a minute tubule which has in its primitive form a ciliated funnel leading from the coelom. At their lateral ends the series of tubules unite to form a duct which grows back to empty into the cloaca. The tubule is associated with a knot of blood vessels near the coelomic opening (the nephrostome). In a more advanced stage of development excretory tubules lack the nephrostome and have the wall expanded to form Bowman's capsule, embracing the knot of blood vessels which is called a glomerulus. This unit, known as a renal corpuscle, is found in the kidneys of most vertebrates. The tubule leading from it is also specialized for the removal of wastes from the blood.

Three pairs of kidneys are found in different vertebrates and appear in succession in the embryos of the higher classes, the reptiles, birds, and mammals. In cyclostomes and embryos of fishes and amphibians the kidney are pronephroi, lying well forward in the body and made up of tubules of the primitive type. The pronephroi are vestigial in embryonic reptiles, birds and mammals. Functional kidneys in these embryos and in the adults of cyclostomes, fishes and amphibia are the mesonephroi, lying behind the pronephroi and made up of closed tubules. As they develop, these tubules connect with the duct formed by the pronephroi; this duct is then called the mesonephric or Wolffian duct. In adult reptiles, birds and mammals the mesonephroi are replaced by the metanephroi, lying still farther back. Their tubules develop in a mass of tissue surrounding a blind diverticulum of the mesonephric duct.

The connection of the excretory ducts with the cloaca persists in many vertebrates but in the true mammals this passage splits to form a dorsal rectum and a ventral urogenital sinus which receives the Wolffian duct. An expanded reservoir, the urinary bladder, developed ventrally in connection with the cloaca, ultimately receives the ducts of the metanephroi, while the remainder of the mesonephric duct persists in the male as the main duct of the testis. The relations of all these parts differ greatly in animals of different groups. In all animals with metanephroi the ducts leading from the kidneys are called the ureters and a separate duct from the urinary bladder to the exterior is the urethra.

EXIT SLIT. A narrow opening in an opaque screen; when a spectrum is produced upon the screen, by a spectrometer, the slit passes only a small portion of the spectrum, which is then focused onto a detector.

EXOGENETIC. A general term designating all surficial, or near surficial, geologic processes such as: erosion, deposition, and secondary enrichment of ore bodies. The term is not particularly applicable to volcanism.

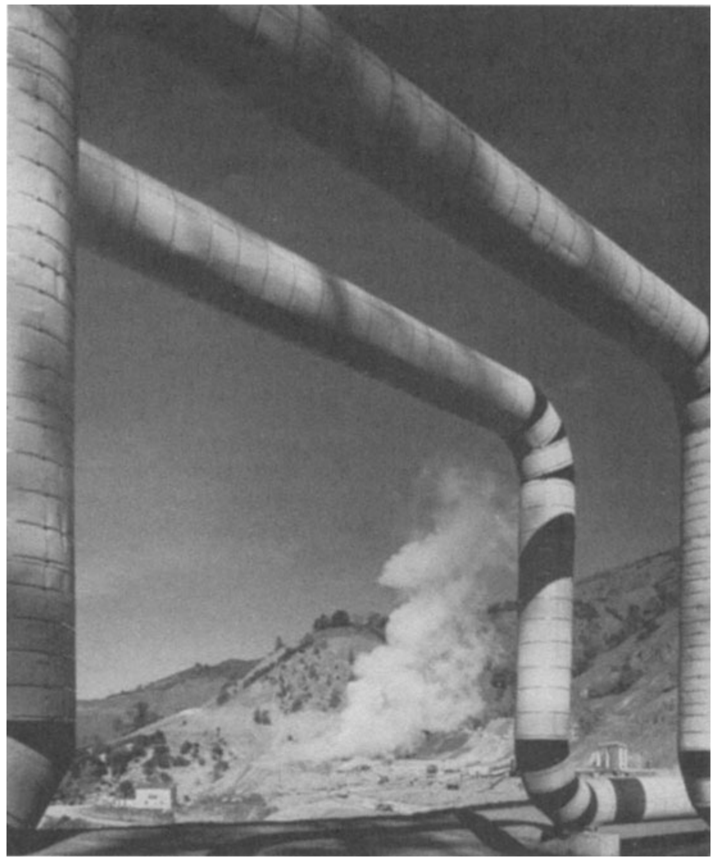
EXOSMOSIS. An osmotic process by which a diffusible substance passes from the inner or closed, to the outer parts of a system, as in the loss of substances from a portion of a plant root to water in the surrounding soil.

EXPANDER (Signal). That part of the communication circuit designed to expand the volume range back to the original value as it is normally compressed for transmission. Weak signals are attenuated and strong signals are amplified.

EXPANSION JOINT. Metals constituting pipes have the property possessed by all materials of expanding with increase of temperature. Were they constrained to a fixed length, a reaction equivalent to the force required to compress the pipe through a deformation equal to the prevented expansion would be set up. For all but very short steam lines this force is too large to incorporate in the piping system. The same force would be present, theoretically, in the short line, but the supports would have enough elasticity to take the small expansion. In long lines the expansion is permitted by the use of suitable joints and bends. See accompanying figure.

Both packed and packless expansion joints are used for saturated steam at pressures up to 250 psi (17 atmospheres). High temperature has a deteriorating effect on packing; however, packed joints have been designed for high temperature by protecting the packing by air-cooled sleeves. Expansion joints take up expansion at one point by allowing relative motion of the two sections of pipe connected by the joint. Usually one pipe end is anchored by a rigid connection to the body of the joint but occasionally the double slip joint in which both pipe ends are free to move in the joint is used.

When expansion is to be taken by the flexibility of the pipe itself various forms of pipe bends are used. This way of caring for expansion is free of the temperature-pressure limitations of the expansion joints and also of any maintenance work such as the repacking of joints. Consequently, it has been the standard for boiler and turbine leads and for long runs of high-pressure piping of all sorts. Its principal drawbacks are the added friction losses, the expense of fabrication (most bends are special jobs), and the space required.



Expansion loops that carry geothermal steam from wells to electric generating station. Located at *The Geysers*, Sonoma County, California. (*Pacific Gas and Electric Co.*)

EXPANSION TURBINE. See **Gas and Expansion Turbines.**

EXPECTED VALUE. The expected value of a function of variate values is its mean value in repeated sampling. If $F(x)$ is the cumulative distribution function of a variate x, the expected value of a quantity t depending on x is

$$\int_{-\infty}^{\infty} t\, dF(x)$$

It is in fact the average or mean value of t over the distribution of x.

EXPLOSIVE. In the conventional sense, a solid, gas, or liquid material which, when triggered, will release a great amount of heat and pressure by way of a very rapid, self-sustaining exothermic decomposition. This entry does not describe nuclear explosives.

There are two principal classes of explosives: (1) *deflagrating explosives* whose burning processes are rather slow—with progressive reaction rates and buildup of pressure that create a heaving action; and (2) *detonating explosives*, which are characterized by very rapid chemical reactions, thus causing tremendously high pressure and brisance (shattering action). In the latter, detonation waves may obtain a velocity in excess of 20,000 feet per second. The decomposition of cellulose nitrate used in propellants typifies the deflagrating type: $C_{24}H_{30}N_{10}O_{40} \rightarrow 5N_2 + 10H_2 + 5H_2O + 11CO_2 + 13CO$. The decomposition of nitroglycerine typifies the detonating type: $4C_3H_5(ONO_2)_3 \rightarrow 12CO_2 + 10H_2O + 6N_2 + O_2$.

Black powder, using KNO_3 or $NaNO_3$, charcoal and sulfur was probably the first explosive developed and is attributed either to Chinese or Egyptian ingenuity. The time of first use occurred before the birth of Christ. This is a deflagrating explosive and was adapted for blasting purposes as early as the 1600s.

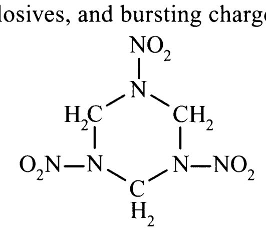
Black powder (gunpowder) consists of an intimate mixture of finely divided solids, 75% potassium nitrate, 15% carbon, 10% sulfur. Powders for sporting guns contain a slightly larger percentage of potassium nitrate (75 to 78%, smaller percentage of carbon (15 to 12%), and a variation in sulfur from 9 to 12%. Mining or blasting powders, where large volumes of gas are desired, may have 14 to 21% carbon and 13 to 18% sulfur. When ignited, potassium nitrate supplies oxygen for the combustion of explosives, of carbon to carbon dioxide and of sulfur to sulfur dioxide. One gram of powder yields 250 to 300 milliliters of gas measured at 0°C and 760 mm pressure. The heat evolved per gram is 500 to 700 calories, and the temperature of the explosion is estimated at 2,700°C.

Among the earliest high explosives were *mercury fulminate*, $HgC_2N_2O_2$, developed late in the seventeenth century, and *nitrostarch*, $C_{12}H_5(ONO_2)_{30}$, discovered by Braconnot in 1832 and still used as a sensitizing ingredient in modern commercial explosives. *Nitrocotton* was produced in 1838 by Dumas and Pelouse by treating cotton and paper with nitric acid, in the same way that Braconnot treated starch with HNO_3. *Nitroglycerin*, $C_3H_5(ONO_2)_3$, was first made by Sobrero. Early nitroglycerin formulations were highly dangerous and caused numerous accidents. In 1867, Nobel found that nitroglycerin could be rendered safe by absorbing it in a porous material, such as kieselguhr, or diatomaceous earth. After formulation of this first *dynamite*, Nobel introduced $NaNO_3$ and later NH_4NO_3 into his dynamite formulations. Nobel, in 1875, while experimenting with cellulose tetranitrate, mixed collodion with nitroglycerine, resulting in the development of blasting gelatin. The development of the blasting cap used with a safety fuse allowed for safe, positive initiation of dynamite.

Wilbrand, in 1863, first prepared *trinitrotoluene* (TNT), $C_6H_2(CH_3)(NO_2)_3$. The material was not manufactured in production quantities until about 1900. The German military recognized the advantages of TNT as a replacement for *picric acid* which they had used earlier. TNT was used extensively during World War I and became a standard military explosive.

Tollens, in 1891, prepared *pentaerythritol tetranitrate* (PETN), $C(CH_2NO_3)_4$, but this compound was not commercially available until after World War I. Commercial production had to await a lowering in the cost of formaldehyde and acetaldehyde used in its production.

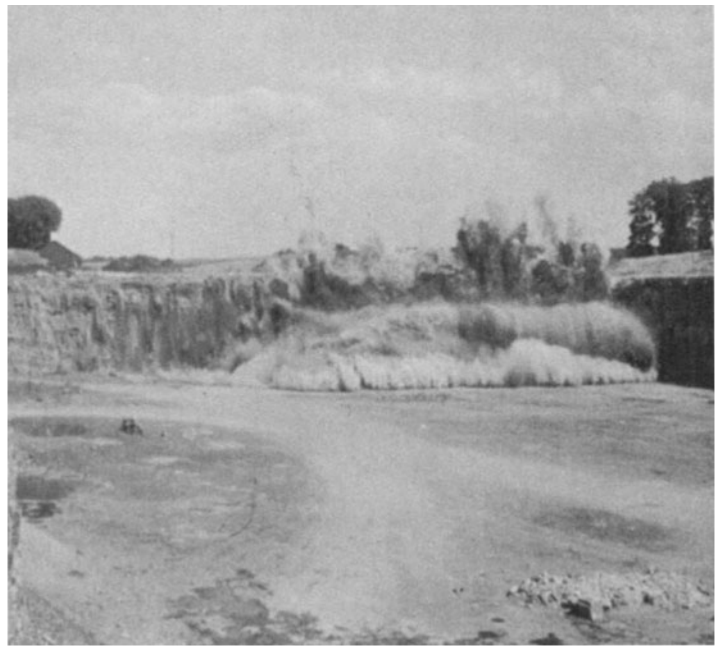
Henning, in 1899, discovered *cyclotrimethylenetrinitramine* (cyclonite-RDX), but its potential was not realized until about 1920. RDX was used extensively during World War II as a component of numerous cyclotols, plastic explosives, and bursting charges.



(RDX)

Although ammonium nitrate, NH_4NO_3, was known to have explosive qualities as early as 1659, it was not used much in explosive formulations until about 1867. At that time, it was used by Nobel to replace a portion of nitroglycerin in dynamite. Because of critical toluene shortages during World War I, NH_4NO_3 was used as a way of conserving TNT. Mixtures containing 80% NH_4NO_3 and 20% TNT; or 50–50 mixes were used. These became known as 80:20 or 50:50 *amatols* and were used as military explosives for shells and bombs. In World War II, particularly by the Axis powers, amatols were used, again to conserve TNT. The Texas City, Texas disaster of 1947 (explosion of ship loaded with NH_4NO_3) reemphasized the potential of this substance for explosives. Later, Robert Akre used a combination of prilled NH_4NO_3 and carbon black (94:6 mix) in making an explosive for blasting in open-pit strip mines. The substance was patented under the name *Akremite*. Later experiments included mixing liquid hydrocarbons to replace the carbon black. This resulted in ANFO explosives. It was found that diesel fuel oil can be mixed with NH_4NO_3 with a consistent quality. Commencing in the late 1950s, ANFO explosives became widely accepted.

Unfortunately, the water resistance of ANFO is low and numerous experiments in attempting to dry-package it were not markedly successful. This shortcoming of ANFO led Cook, Farnum, and others to develop *slurry explosives*. These materals are comprised of oxidizers, such as NH_4NO_3 and $NaNO_3$, fuels, such as coals, oils, aluminum, and other carbonaceous materials, sensitizers, such as TNT, nitrostarch and smokeless powder, and water—all mixed with a gelling agent to form a thick, viscous explosive having excellent water-resistant properties. These explosives are made as cartridged units or are mixed at the site in bulk and then pumped into place.



Quarry blast utilizing high explosives initiated with electrical delay devices (millisecond delay electric blasting caps).

The largest consumers of commercial explosives are the mining, construction, and seismic-prospecting industries. Explosives are also used in agriculture for blasting stumps, setting posts, breaking up hardpan, clearing land, digging wells, and blasting drainage and irrigation ditches. Explosive technology also is applied by industry in the forming, cladding, bonding, hardening, and welding of metals. On a small scale, explosive-actuated devices are used in valves, switches, and relays, as well as for cutting, punching, riveting, and fastening metals.

See accompanying illustration of a typical quarry blast using a millisecond delay electric blasting cap.

EXPONENT. Also called index, it is the power to which an algebraic expression is raised. It can be positive or negative, integral or fractional. If m and n are positive integers and a, b are numbers or functions, the following are some of the properties of the exponent:

$$
\begin{aligned}
a^m \times a^n &= a^{(m+n)} \\
a^m/a^n &= a^{(m-n)} \\
a^{-m} &= 1/a^m \\
(a^m)^n &= a^{mn} \\
a^{1/m} &= \sqrt[m]{a} \\
\sqrt[m]{\sqrt[n]{a}} &= \sqrt[mn]{a} \\
a^{m/n} &= \sqrt[n]{a^m} \\
(ab)^m &= a^m b^m \\
(a/b)^m &= a^m/b^m \\
\sqrt[m]{ab} &= \sqrt[m]{a}\,\sqrt[m]{b} \\
\sqrt[m]{a/b} &= \sqrt[m]{a}/\sqrt[m]{b}
\end{aligned}
$$

An exponential function is transcendental, an example being $y = ab^x$, where a, b are constants and x is the independent variable. If a

= 1, it is the inverse of the logarithmic function. Some properties of its curve are: (a) it is not symmetric to either the X- or Y-axis or to the origin; (b) it intersects the Y-axis at $y = 1$ but its asymptote is the X-axis; (c) no finite value of x makes y infinite. If $b < 1$, $y \to 0$ as $x \to \infty$ and decreases continuously as x increases; if $b > 1$, $y \to 0$ as $x \to -\infty$ and increases continuously with x; if $b = 1$, the curve becomes the straight line $y = 1$.

The most convenient value to choose for b is the transcendental number e. Its curve is similar to the more general one described for the number b but its slope is different.

An exponential equation contains one or more exponential functions. It can often be solved by taking the logarithm of both sides and solving the resulting algebraic equation.

Euler's theorem on the exponential function is $\cos x \pm i \sin x = e^{\pm ix}$.

EXPONENTIAL DISTRIBUTION. A distribution of the form

$$dF = \frac{1}{\sigma} \exp\left(-\frac{x-m}{\sigma}\right) dx, \quad m \le x \le \infty$$

The parameter σ is the standard deviation of the distribution and is equal to the distance of the mean from the start.

EXPONENTIAL SMOOTHING. A method used in forecasting a variable from values of that variable occurring at previous points of time. It relies on a weighted average of those previous values, and on the reasonable assumption that values in the more remote past are of less influence on the present, the weights attached to the values diminish for the less recent values. Such a forecast at time t of a variable u_t, for example, might be

$$u_t(\text{forecast}) = (1 - \beta) \sum_{i=1}^{\infty} \beta^{j-1} u_{t-j}$$

The coefficients diminish according to the exponent of the coefficient β; hence the name of the procedure. More elaborate forms of the same basic method are also known as exponential smoothing.

EXPRESSION (Mechanical). The separation of liquids from solids by compressively squeezing certain liquid-containing substances, such as separating oils from vegetable seeds and nuts. Equipment is designed to permit the liquids to be removed while still retaining the solids between the compressing surfaces. Expression equipment takes several configurations. In the *plate press*, the material to be expressed, such as fruit or seeds, is wrapped in special plate cloths. These then are placed between a series of hydraulically operated plates, which act upon the materials much as a vise. The pressure usually is applied in stages, with the maximum pressure applied close to the end of a 20-to-45-minute total cycle. The *box press* is similar with the exception that shallow boxes enclose the pressed cake on two sides, thus simplifying the folding of the press cloths. The *cage press* consists of a cylinder, finely perforated, with a hydraulically-operated ram. Essentially, the material to be expressed is squeezed at one end of the cylinder by the ram. The expressed oil flows through the perforations. Several pressings (strokes of the piston or platen) are usually required. In the *pot press*, the cage is replaced by a series of short, superimposed steam-heated pots. Usually a series of pots is used in each press, the bottom of each pot serving as the ram for the pot below. The *curb press* also is similar to the cage press, but operates at lower pressures with fewer drainage channels. The *screw* press consists of a continuous screw or worm that rotates within a cyclinder housing lined with perforated plates. A powerful squeezing action results from the taper of the screw. In the *V-disk press*, two conical disks face each other in a suitable casing. As the disks rotate, they converge from a point of maximum gap (point of feed) to a point of minimum gap (point of discharge). Designs with hydraulic systems permit the disks to oscillate during operation to maintain constant pressure. The *roll press* operates on the principle of the old-fashioned clothes wringer, with two to three rollers through which the feed is passed, the expressed liquids collecting in a trough below.

Operating characteristics of the various forms of expression equipment are summarized in the accompanying table.

EXPRESSION EQUIPMENT

Type of Press and Applications	Oil, Fat, Liquids in Feed	Cake Discharge
Plate. Vegetable seeds, nuts, olives, fruit, notably flaxseed	30–35%	5–10% oil
Box. Vegetable seeds, nuts, notably peanuts and cottonseed	30–35%	5–10% oil
Cage. Most oil seeds and nuts. Almost any oil material, notably copra and castor beans	35–50%	5–10% oil
Pot. Fats, such as cocoa butter not liquid at room temperature	30–50%	6–10% fat
Screw (Low-pressure type). Wastewater slurries, wood and wood pulp; food and beverage products	90–97%	25–50% solid
Screw (High-pressure type). Vegetable seeds and nuts; rubber; rendered materials	30–35%	3–5% oil
V-disk. High-polymer resins, spent grains, wood and pulp products, food and starch products, wastewater slurry	85–97%	25–55% solid
Roll. Alkali cellulose, sugar cane, wood and pulp products	92–98%	30–55% solid

EXSOLUTION. See **Mineralogy.**

EXTENSOMETER. A device for determining small linear dimensional changes caused by the application of a stress. Thus, an extensometer is used to determine the changes in length of the gage section of a tensile specimen during a tensile test.

EXTINCTION COEFFICIENT. 1. A measure of the space rate of diminution, or extinction, of any transmitted light; thus, it is the attenuation coefficient applied to visible radiation. The extinction coefficient is identified in a form of Bouguer's law (or Beer's law):

$$dI = -\sigma I \, dx$$

or

$$I = I_0 e^{-\sigma x}$$

where I is the illuminance (luminous flux density) at the selected point in space, I_0 is the illuminance at the light source, and x is the distance from the source.

When so used, the extinction coefficient equals the sum of the medium's absorption coefficient and scattering coefficient, each computed as a weighted average over all wavelengths in the visible spectrum. So long as scattering effects are primary, as in the lower atmosphere, the value of the extinction coefficient is a function of the particle size of atmospheric suspensoids. It varies in order of magnitude from 10 km^{-1} with very low visibility to 0.01 km^{-1} in very clear air.

The extinction coefficient is related to the transmission coefficient τ as follows:

$$\tau = e^{-\sigma}$$

2. In oceanography, the extinction coefficient is a measure of the attenuation of downward-directed radiation in the sea. The coefficient K is defined by

$$K = 2.303 \log \frac{I_{\lambda 1}}{I_{\lambda 2}}$$

where $I_{\lambda 1}$ is the intensity of radiation of a given wavelength λ on a horizontal surface and $I_{\lambda 2}$ is the intensity on a horizontal surface 1 meter deeper. K varies with wavelength, with the nature of the scattering particles, and with the presence of dissolved colored substances.

EXTRACTION (Liquid-Liquid). Sometimes referred to as solvent extraction, this operation is effected by treating a mixture of different substances with a selective liquid solvent. At least one of the components of the mixture must be immiscible or partly miscible with the treating solvent so that at least two phases can be formed over the entire range of operating conditions. To be effective, one or more of the components must be dissolved from the mixture by the solvent preferential to the other components present. The solvent-rich phase that contains the preferentially dissolved component is termed the *extract layer*. The residual phase formed by the undissolved component (or diluent) and usually containing some solvent is called the *raffinate layer*. Either layer may be at the top or bottom of the separating vessel, depending upon relative densities. Other forms of solvent extraction include leaching, washing, and precipitative extraction (*salting out*).

Liquid-liquid extraction finds application in separating the components of condensed mixtures where vaporization methods, such as distillation or evaporation, may be impractical. This condition may arise because the substances to be separated may have comparable volatilities, are relatively nonvolatile, are heat-sensitive, or have one component present in very small concentration.

Liquid-liquid extraction finds wide application throughout the processing industries, including the manufacture of toluene, uranium vanadium, amino acids, coal-tar products, lube-oil refining, protein processing, and solvent refining of coal and oil shales.

Solvent extraction also is an important laboratory operation as in the recovery of oils from oil-bearing material. The material is placed in a porous container and subjected to treatment with solvent. The solvent containing some dissolved material passes through the porous membrane, leaving the undissolved residue in the container. The principle of counter current extraction may be utilized in consecutive containers, or the solvent may be vaporized from the solution, condensed onto the material and, by means of a syphon in the apparatus, withdrawn periodically to the solution compartment below, as in the Soxhlet type of apparatus. When a third substance is of different solubility in two nonmiscible liquids, this substance may be separated from the solution of lower concentration by shaking with the preferential solvent, and then separating the two liquid layers. The desired substance may be recovered from the solution by evaporation of the solvent. The effectiveness of separation is increased by the use of a given amount of extracting solvent in successively smaller portions rather than by a single extraction with the total amount.

Example: upon shaking 1 volume of liquid A plus 1 volume of liquid B, assume a concentration ratio of 1 (concentration in B)/10 (concentration in A) of the third substance C. See Table 1. Upon shaking 1 volume of liquid A plus $\frac{1}{2}$ volume of liquid B and, after separation, shaking 1 volume of liquid A (containing the residue of C) plus $\frac{1}{2}$ volume of liquid B, the results are indicated by Table 2.

A single equal-volume extraction would, therefore, remove 91% of C from A, whereas a double half-volume extraction would remove 97%.

TABLE 1. LIQUID-LIQUID EXTRACTION USING EQUAL VOLUMES IN SINGLE EXTRACTION

Concentration Ratio $\frac{B}{A}$	Volume Ratio $\frac{B}{A}$	Amount of C in		Fraction of C in	
		B	A	B	A
10	1	$10 \times 1 = 10$	$1 \times 1 = 1$	$\frac{10}{11} = 0.91$	$\frac{1}{11} = 0.09$

TABLE 2. EFFECT OF SUCCESSIVELY SMALLER PORTIONS ON LIQUID-LIQUID EXTRACTION

	Concentration Ratio $\frac{B}{A}$	Volume Ratio $\frac{B}{A}$	Amount of C in		Fraction of C in	
			B	A	B	A
First extraction	10	0.5	$10 \times 0.5 = 5$	$1 \times 1 = 1$	$\frac{5}{6} = 0.83$	$\frac{1}{6} = 0.17$
Second extraction	10	0.5	0.17×0.83	0.03	0.14	0.03
Combined					0.97	0.03

EXTRACTIVE DISTILLATION. See **Distillation.**

EXTRACTIVE METALLURGY. That phase of metallurgy dealing with the removal of metals from minerals. Methods are discussed under individual metals.

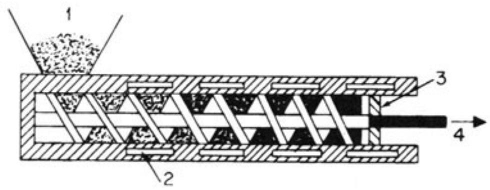
EXTRAEMBRYONIC MEMBRANES. A series of structures developed in connection with the embryos of vertebrates (reptiles, birds and mammals), but not as parts of the body itself. They relate the embryo to its environment in several ways. These membranes are the allantois, amnion, chorion, serosa, and yolk sac.

EXTRAORDINARY INDEX. The refractive index for the extraordinary ray in a crystal showing double refraction, measured perpendicular to the optic axis (in which direction its value differs most from the index for the ordinary component). If an unpolarized ray of light strikes the surface of calcite or other crystal normally showing double refraction it will be divided into two transmitted rays. One ray, the ordinary, will not be bent, while the other ray, the extraordinary, will be bent on entering the crystal. Rotation of the crystal about the entrant ray causes the extraordinary ray to rotate about the ordinary ray.

EXTRUSION. A majority of stock plastic shapes (bars, cylinders, special cross sections) are made in this way. Thermoplastic materials are heated in a plasticizing cylinder and by means of a rotating screw are forced through a die to provide the desired cross section. A variation of the process is used for extruding coatings of soft plastic materials over other materials. Almost any profile can be imparted to the product, but of course, variations in profile are limited to two dimensions. The tooling costs for extrusion are low compared with injection molding. Thickness of the material can be controlled quite precisely. Production rates are high.

Certain metals, including aluminum, and various rubbers are also extruded.

Extrusion is also combined with mixing in some applications. In the type of device shown in the accompanying diagram, the material is fed as a dry solid, is fluxed in the barrel to form a paste, and then resolidified at the discharge. Action in the barrel is one of shearing, rubbing, and kneading. One continuous screw or two screws rotate in the closely-fitting barrel. The work of the screw is augmented by forcing the material through breaker screens and around breaker disks just before the material is forced out through the exit nozzle. Such machines are often used for extruding soft chemical and food mixes which do not require fluxing, as well as for the extrusion of hard plastics, some of which must be fluxed at temperatures above 400°F (204°C). Wires can be covered and shapes of intricate cross section can be produced. Plastic resins also are blended in extruders to form pellets for later press and injection molding.



Mixed-type extruder: (1) Charge stock; (2) heating or cooling chambers in extruder jacket; (3) die; and (4) extruded product.

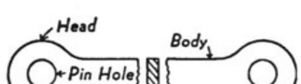
An excellent and detailed summary of the use of extrusion in the plastics and polymer industries is given by J. A. Gibbons, et al. in the "Modern Plastics Encyclopedia," pp. 219-234, McGraw-Hill, New York, 1986.

Extrusion in the Food Industry. Numerous food formulations, as exemplified by spaghetti, macaroni, and other pasta products, are characterized by a uniform cross-sectional shape (round, rectangular, etc.) and length (rodlike, tubelike, ribbonlike, etc.). These characteristics are imparted by forcing an initial pasty mass through the dies of an extruder. Temperature control of extruding is extremely important. For more detail, reference is suggested to the "Foods and Food Production Encyclopedia," (D. M. and G. D. Considine, Eds.), Van Nostrand Reinhold, New York, 1982.

EXTRUSIVE. A property of an igneous rock that has been ejected onto the earth's surface. Lava flows and detrital material, such as volcanic ash, are extrusives. See also **Mineralogy.**

EYE BAR. A heat-treated tension member formed from a single piece of steel. The finished eye bar consists of a body having a rectangular cross section and two circular heads containing holes for pins, which are used to connect the eye bar when it forms part of a structure.

In the fabrication of an eye bar, the ends of a steel plate, of the correct cross-sectional area and length, are heated and upset to form the heads. The heads are next rolled to remove any unevenness resulting from the upsetting operation. While the ends are still hot, holes are punched out which are smaller in diameter than the finished pin holes. The bars are then subjected to special heat treatment which produces a high tensile strength. After cooling, the pin holes are bored to exact size simultaneously.



Eye bar.

Eye bars make excellent tension members since the heat treatment enables them to carry higher tensile loads than the ordinary built-up steel members. As the eye bar is a very slender member, it cannot be used where there is a possibility that it will have to carry compressive stress. These members are used in the cable anchorages of suspension bridges. Eye bar chains are sometimes used instead of wire cable for suspension bridges. In the past, eye bars were extensively used for tension members of truss bridges. The use of riveted joints instead of pin-connected joints made eye bars obsolete for such use.

EYE (Human). See **Vision and the Eye**

EYELID. A fold of skin which can be drawn over the eye in vertebrates above the fishes. Three eyelids are the maximum. These are an upper and lower lid and a third eyelid or nictitating membrane which passes between the others and the eyeball from the inner to the outer margin of the eye. The eyelids contain glands whose secretion lubricates the apposed surfaces of the lids and eyeball, and in the mammals bears a row of stiff hairs, the cilia or eyelashes.

In humans, *hordeolum* (sty) is a common infection of one or more of the small glands of the eyelids, usually caused by staphylococci. Children are especially susceptible. A sty begins as a small, reddened area on the margin of the lid. Pain is almost always present and is directly related to the amount of swelling. In severe cases, the entire eyelid is swollen. A few days after its appearance, the sty develops a yellow center, caused by the formation of pus, and usually erupts a few days later. A single sty may not require medical attention unless it is quite painful. When a number of sties appear, or when they recur often, general health and diet should be evaluated.

Chalazion is a swelling or enlargement of one of the oil glands of the eyelid. This is caused by obstruction of the gland's duct. The skin moves loosely over the swelling. The physician may prescribe topical medication, but, if this fails, a simple surgical procedure is performed which removes the mass, leaving no visible scar.

Blepharitis is a relatively common condition in which the margins of both eyelids become red and inflamed. Blepharitis can be caused by bacterial infection or it may be an extension of seborrheic dermatitis, involving the scalp, eyebrows, and, at times, the ears. Blepharitis may occur only as redness with slight crusting, or it may cause itching, burning, and edema of the eyelids, lacrimation, and hypersensitivity to light. The lids often become stuck together overnight from the accumulation of dried secretions.

Ptosis is a condition in which one or both upper eyelids droop. This is caused by the failure of the levator muscles of the eyelid to operate properly. The abnormality may be congenital or acquired. Congenital ptosis, when severe, may be treated by surgical alteration of the involved muscles.

Edema of the eyelids usually results from allergies to eyedrops, drugs, or cosmetics. Trichinosis also may produce eyelid edema. See **Trichinosis.**

See also **Vision and the Eye.**

EYEPIECE. Also known as the *ocular*, the lens, or system of lenses, closest to the eye in an optical instrument such as a telescope or a microscope. The eyepiece is usually a magnifying device used for the purpose of detailed examination of the real image formed by the objective of the instrument. It is usually designed to act as a collimator to the light from the objective, so that the light from each point of the image formed by the objective emerges in parallel or nearly parallel rays. Hence, in using a telescope or microscope in proper adjustment, the eye should be focused as though looking at a distant object.

The simplest type of eyepiece is either a simple convex or concave lens, of relatively short focus, so placed as to serve as a magnifier for the image formed by the objective. Because of the spherical and chromatic aberrations of the simple lens of short focus, a combination of lens is usually employed as an eyepiece. The two most common types of compound eyepieces are the Huygens (Fig. 1) and the Ramsden (Fig. 2). In these eyepieces, the lens F is known as the field lens and the lens E as the eye lens. The Huygens eyepiece is placed slightly inside the focus of the objective, and the field lens of the eyepiece forms a real

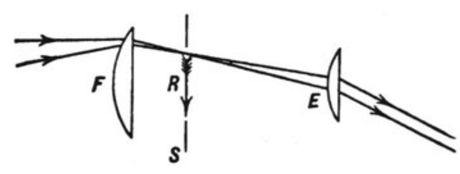


Fig. 1. Huygens eyepiece.

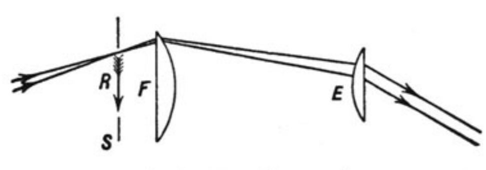


Fig. 2. Ramsden eyepiece.

image R in the plane S, from which the rays emerge parallel from E. In the Ramsden type, the field and eye lenses combine to render the light from the real image R, formed in the plane S by the objective, parallel upon emergence from the eyepiece. Since the Huygens type eyepiece is placed inside the principal focus of the objective, a reticle or filar micrometer cannot be used, although a reticle may be placed inside the eyepiece itself in the plane S. The Ramsden type, on the other hand, is focused directly upon the plane of the real image from the objective, and a reticle or micrometer may be placed in this plane. Eyepieces of the Ramsden type, which are simple magnifiers focused upon the real image from the objective, are known as positive eyepieces; while eyepieces placed inside the principal focus of the objective, as in the case of the Huygens type, are known as negative eyepieces. (See entries on individual optical instruments for further discussion of positive and negative eyepieces.)

Both the positive and negative eyepieces give a view of the image from the objective in the same orientation as that image is formed. This means that the observer will see the image of a distant object inverted. Although this is no disadvantage in microscopes and in astronomical telescopes, it is intolerable in a telescope or field glass to be used for observation of distant terrestrial objects. The simple concave lens, as used in the so-called Galilean telescope or opera glass, gives an erect image of a distant object. To avoid the aberrations of the simple concave lens, various "erecting systems" are used in terrestrial telescopes. Some of these erecting systems employ prisms, as in the case of binoculars, or complicated systems of lenses.

EYE (Vertebrate). A sensory organ which is stimulated by light, particularly an organ whose stimulation results in the formation of a mental image of the objects from which the light is reflected or radiated.

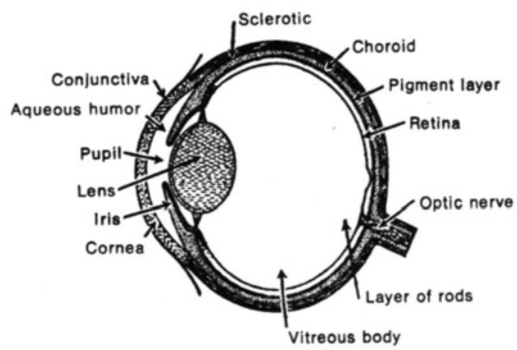
Details of the human eye are described under **Vision and the Eye.**

Although most eyes enable the animal to form visual images, that is to see in the usual sense of the word, some light-sensitive organs are capable only of perceiving light and the direction from which it comes. Pigment spots in some of the 1-celled animals are supposed to be light-sensitive and in flatworms and a few insects the eyes are formed of a group of sensitive cells partially isolated by pigment. The structure of these eyes shows no possibility of their forming images.

Eyes of reasonable complexity are found in some of the segmented worms and three types of complex eyes occur in the phyla *Mollusca, Arthropoda*, and *Chordata*. These three forms of eyes have been extensively studied and are known in detail.

Arthropod eyes are of two kinds, simple and compound, of which one or both may occur in a single individual. The sensory end organ in both forms is the retinula, a group of visual cells surrounding a central optical rod or rhabdom. In many simple eyes a portion of the cuticula is thickened to form a biconvex lens opposite to a group of retinulae. Compound eyes are made up of many ommatidia, each consisting of a similar lens forming a facet of the cornea of the entire eye, and an underlying retinula, with intervening crystalline cells and, in some species, other structures.

Both mollusks and vertebrates, including, of course, humans, have camera eyes, although their development and structure differ. All camera eyes have a lens suspended before a chamber lined with a sensory layer, the retina. In front of the lens is another chamber and in front of that a transparent cornea which acts as a lens in terrestrial animals. The eye is insulated by a heavily pigmented layer which surrounds it except where the lens is suspended and extends in front of the lens as the iris. The iris, activated by muscles, controls the size of its central opening, the pupil, through which light enters the eye. Light passing through the lens is focused on the retina in a sharp image and the varied stimuli acting on nerve endings result in a definite mental picture. Such eyes are provided with muscles which direct them toward objects to be observed. They also have muscular focusing devices which move the lens in relation to the retina or vice versa, or control the curvature of the lens as in the human eye.



The vertebrate eye.

The action of the different kinds of eyes results in different kinds of vision.

For vision in fishes, see **Fishes;** in snakes, see **Snakes.**