N

NADIR. The point exactly opposite the zenith on the celestial sphere is termed the *nadir*. See also **Celestial Sphere and Astronomical Triangle.**

NAIAD. An aquatic insect larva.

NAND (Circuit). A computer logical decision element which has the characteristic that the output F is 0 if, and only if, all the outputs are 1's. Conversely, if any of the input signals A or B or C of the three-input NAND element shown in Fig. 1 is not a 1, the output F is a binary 1. Although the NAND function can be achieved by inverting the output of an AND circuit, the specific NAND circuit requires fewer circuit elements. A two-input transistor NAND circuit is shown in Fig. 2. The output F is negative only when both transistors are cut off. This occurs when both inputs are positive. The number of inputs, or fan-in, is a function of the components and circuit design. See also **AND (Circuit).** NAND is a contraction of NOT AND.

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

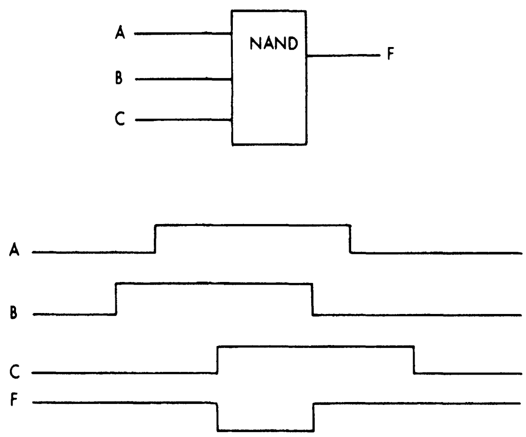


Fig. 1. Schematic of a NAND circuit.

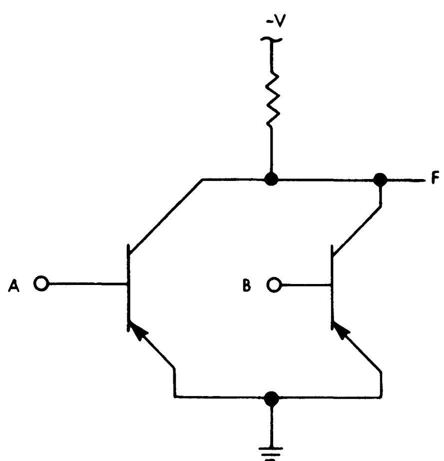


Fig. 2. Transistor-type NAND circuit.

NANNOPLANKTON. The portion of the floating and drifting aquatic animals (plankton), including minute species, which pass through ordinary nets and must be secured by centrifuging.

NANSEN BOTTLE. A bottle used for collecting samples of seawater at any desired depth. After the bottle has been lowered to the desired depth, a weight is sent sliding down the wire to which the bottle is attached. This releases a catch when it strikes the bottle which in turn closes the valves and traps the water inside. Special thermometers may be used to record the temperature at the desired depth.

NAPHTHALENE. See **Coal Tar and Derivatives; Organic Chemistry.**

NAPIER'S RULES. See **Spherical Trigonometry.**

NARWHAL. See **Whales, Dolphins, and Porpoises.**

NATIONAL BUREAU OF STANDARDS (U.S.). The NBS was established by the U.S. Congress in 1901. The name of the institution was changed just a few years ago to the National Institute of Standards and Technology (NIST). See **NIST.**

NATIONAL CENTER FOR ATMOSPHERIC RESEARCH (NCAR). This organization was founded in 1960 by a consortium of 14 universities with doctoral programs in atmospheric sciences. Its staff numbers about 750 people. The headquarters is in Boulder, Colorado. By the early 1990s, the consortium had grown to more than 50 members in the United States and Canada. NCAR's chief source of funding is the U.S. National Science Foundation.

NCAR's research consists of four basic areas: (1) atmospheric chemistry; (2) climate and its links with other environmental systems; (3) solar and solar-terrestrial physics; and (4) microscale and mesoscale meteorology (the study of phenomena as small as the formation of ice crystals in clouds and as large as major thunderstorm systems). Additionally, one group of researchers looks at how societies and individuals respond to changes in their environment. All of the foregoing activities involve field and laboratory studies, theoretical work, and computer modeling.

Supercomputing power and observing facilities are among the services provided by NCAR because they are too costly and specialized for individual universities to acquire. NCAR also adapts research technology for practical applications, particularly in the area of aviation safety. In 1990, NCAR acquired a CRAYY-MPS/864 supercomputer, which, at the time, was the most powerful computing system in existence. With such computer power, researchers are able to construct models, or mock worlds, allowing them to explore speculative scenarios of such events as the fates of chemicals in the atmosphere, the formation of clouds, the global circulation of winds, and the movements of ocean eddies. These models are important for exploring phenomena that cannot be observed or studied firsthand.

Much of the information for NCAR's projects is gained from advanced observing systems. Included in such systems are advanced radar technology based on the Doppler effect and a fleet of research aircraft, which carry a variety of sophisticated instruments. Projects directed toward early practical application include improvement of aircraft flight safety and the efficient use of U.S. airspace. Surface wind shifts, microbursts, tornadoes, hail, lightning, and heavy rain pose hazards and compromise the efficiency of the aviation industry. So-called "Nowcasts" are designed to enable predictions within a few minutes to an hour, as contrasted with longer-term hazardous weather forecasts. A program known as Terminal Doppler Weather Radar detects and warns of wind shifts near the airport terminal, approach, and departure zones, and runways.

NATIONAL RESEARCH COUNCIL (NRC). The Council was established by the National Academy of Sciences in the United States in 1916 to associate the broad community of science and technology with the Academy's purpose of furthering knowledge and of advising the federal government. The Council operates in accordance with general policies determined by the Academy under the authority of its congressional charter of 1863, which establishes the Academy as a private, nonprofit, self-governing membership corporation. The Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in the conduct of their services to the government, the public, and the scientific and engineering communities. It is administered jointly by both academies and the Institute of Medicine. The National Academy of Engineering and the Institute of Medicine were established in 1964 and 1970, respectively, under the charter of the National Academy of Sciences.

NATROLITE. The mineral natrolite, one of the zeolites, is a sodium aluminum silicate corresponding to the formula $Na_2Al_2Si_3O_{10} \cdot 2H_2O$. It is orthorhombic, crystallizing in slender prisms of nearly square cross-section which are terminated by relatively flat pyramids. There are also fibrous to compact varieties. Natrolite is a brittle mineral; hardness, 5-5.5; specific gravity, 2.2; luster, vitreous; color, red, yellow, white, or colorless; transparent to opaque. Natrolite is found with other zeolites in fissures and cavities in basaltic and related rocks. Czechoslovakia, France, Italy, Norway, Scotland, Ireland, Iceland, Greenland, and South Africa contain well-known localities for natrolite. In the United States it is found in the Triassic traps of New Jersey; also from Oregon, Washington, Montana, Colorado, and as exceptional crystals from San Benito County, California. Superb crystals occur at Mt. St. Hilaire, Quebec, Canada, and from an asbestos mine in Quebec, crystals up to 3 feet (0.9 meters) long and 4 inches (10 centimeters) in diameter have been found. The name natrolite refers to its soda content.

NATURAL COORDINATES. An orthogonal, or mutually perpendicular, system of curvilinear coordinates (see **Coordinate System**) for the description of fluid motion, consisting of an axis t tangent to the instantaneous velocity vector and an axis n normal to this velocity vector to the left in the horizontal plane, to which a vertically directed axis z may be added for the description of three-dimensional flow.

NATURAL FREQUENCY (Mathematics). A term broadly applied to any system whose transfer function approximates a second-order differential equation of the form $s^2 + 2\zeta\omega_n s + \omega_n^2$, where ω_n is the natural frequency, ζ is the damping ratio, and s is the Laplace transform operator. In control and feedback systems, for example, continuous oscillation or hunting may occur. The frequency of this oscillation is the natural frequency.

NATURAL GAS. A major source of energy for industrial, commercial, and domestic needs, natural gas is consumed by numerous countries worldwide. Because natural gas is comparatively easy to transport over long distances, usage is not confined to regions that produce it. In addition to energy, natural gas also is a critically important source of industrial chemicals, including numerous hydrocarbon-based organics that find ultimate usage in plastics, films, fibers, solvents, and coatings.

In terms of interest as a fossil fuel for generating electrical power and other energy-conversion processes, natural gas is gaining favor. Compared with coal, natural gas frequently is termed the "clean-burning" fuel. As compared with coal as an energy raw material, the "add-on" costs for treating combustion effluents to satisfy environmental requirements are less for natural gas than for coal, and, consequently, the lower cost benefits of coal are eroding. Further, improvements in natural gas production technology and a brighter outlook for natural gas reserves are contributing to the expansion of natural gas consumption. In addition, large advances have been made in the combustion efficiency of natural gas. For example, the efficiency of some domestic heating appliances has increased to about 95% as compared with 60% or 70% efficiency a relatively few years ago. The cogeneration of heat and electricity in industrial utilities is tending to favor natural gas as the raw fuel. In terms of local natural gas distribution, utilities are finding the use of polyethylene pipe an important cost-saving factor.

In a summary of the Gas Research Institute (GRI)[1], the following observation is made: "The United States, the gas industry, and the gas consumer have entered a dynamic new decade filled with change, challenge, and opportunity. These include the reemergence of the environmental movement, the continuing deregulation of the U.S. natural gas market, the expansion of global trade in the former Soviet Bloc, the emergence of more competitive international and national energy markets, and the rapid expansion of technology options. Three stragetic needs are likely to dominate the 1990s for the U.S. gas industry and the gas consumer: (1) ensuring gas deliverability while controlling costs to the consumer; (2) responding to increased concern for the environment; and (3) satisfying a demand for higher quality of energy service."

See also **Electric Power Production and Distribution.**

Composition of Natural Gas

The composition of natural gas varies with the source, but essentially is made up of methane, ethane, propane, and other paraffinic hydrocarbons, along with small amounts of hydrogen sulfide, carbon dioxide, nitrogen, and, in some deposits, helium. Natural gas is found underground at various depths and pressures, as well as in solution with crude-oil deposits. Principal gas deposits are found in the United States, Canada, the former Soviet Bloc, and the Middle East. The analysis of a gas sample taken from the Panhandle natural gas field in Texas is given in Table 1. Because numerous parts of the earth do not have natural gas at all, or where supply is less than demand, much natural gas is transported, notably by pipeline in the gaseous or liquid phase and across the seas in specially-designed LNG (liquefied natural gas) carriers.

TABLE 1. ANALYSIS OF NATURAL GAS FROM NATURAL GAS FIELD IN TEXAS PANHANDLE

Component	Mole Percent
Methane	76.2
Ethane	6.4
Propane	3.8
Normal butane	1.3
Isobutane	0.8
Normal pentane	0.3
Isopentane	0.3
Cyclopentane	0.1
Hexane plus other hydrocarbons	0.35
Nitrogen	9.8
Oxygen	Trace
Argon	Trace
Hydrogen	0.0
Hydrogen sulfide	0.0
Carbon dioxide	0.2
Helium	0.45

NOTE: Heating value of various natural gases averages between 975 and 1180 Btu/cubic foot (8678-10,502 Calories/cubic meter) at 60°F (15.6°C) and 30 inches (76.2 centimeters) mercury pressure.

Origin and Geology of Natural Gas

The most commonly accepted theory concerning the formation of natural gas is the organic theory. Methane is a product of decaying vegetable matter and in areas of stagnant water is found as *marsh gas* or *swamp gas*. It is theorized that over millions of years, the remains of plants and animals were washed down into lakes, the accumulations

[1]"1993–1997 Research & Development Plan," Gas Research Institute, Chicago, Illinois.

covered with layers of mud and stone. The latter became stone while the organic matter decayed through the action of heat and pressure and perhaps from effects of bacteria and radioactivity, forming various hydrocarbons. The hydrocarbons were held in tiny spaces between the particles of sand and porous rock and formed natural gas and petroleum. Often the natural gas so formed made its way through the rock to the surface and escaped. In some areas, however, the layers of sand and porous rock were covered by impermeable rock to form huge reservoirs of natural gas at various levels of pressure. See also **Petroleum.**

The organic theory as usually presented is rather general and vague in many respects. As observed by Ourisson (Université Louis Pasteur, Strasbourg) and colleagues, natural gas, as well as coal and petroleum, are fossil fuels, but fossils of what? Fossil fuels form only if the organic matter is buried before it can become completely oxidized to carbon dioxide by microorganisms. According to the microbial origin concept, as the carbon compounds sink deeper into the Earth under accumulating sediments, they are subjected to high temperatures and undergo chemical reaction, during which oxygen and most other elements are eliminated, thus yielding a mixture composed in the case of gas and petroleum almost entirely of hydrocarbons (carbon in the case of coal). Since the beginnings of photosynthesis on Earth, it is estimated that 10 quadrillion (10^{16}) tons of carbonaceous material has been stored in sediments. Most of this material is stored in very dilute form and only under exceptional geologic conditions, is it concentrated to become a viable fuel source. In a twenty-year study which might be called molecular paleontology, Ourisson and coworkers have been studying the detailed genesis of fossil fuels. Thus far, chemical analysis of the most varied organic sediments reveals a surprising commonality—all appear to derive much of their organic matter from once unknown microbial lipids. This topic is presented in more detail in entry on **Petroleum.**

A few scientists, notably Thomas Gold (Cornell University), have proposed that, in contrast with the organic sediments theory, the prime source of natural gas is primordial, abiotic methane rising from deep within the Earth's mantle. This is sometimes referred to as the "deep-earth gas" hypothesis. In this view, methane flows up around the edges of the shield and is responsible for the oil and gas fields in the North Sea and the southern Baltic. Admittedly, this reservoir of natural gas is presently out of the reach of any foreseeable drilling technology. In some areas, it is suggested that the granite crust may have been fractured and subsequently became porous to the extent that methane may have risen into the crust and have become trapped at accessible depths. Other scientists point out that no evidence supports the concept that a large amount of methane was incorporated in the Earth when it was formed. Available geochemical evidence suggests that the early atmosphere, produced by outgassing of the planetary interior, could not have been rich in hydrogen. Further, if the Earth did at one time contain primordial methane or other hydrocarbons, most of that volatile material would have long since escaped by way of volcanism and diffusion. It is also suggested that the analysis of volcanic basalts shows that the rock in the upper mantle is highly oxidizing, in which case any methane present would have been converted to carbon dioxide.

Some experimental drilling programs underway in Sweden, including a well some 5000 meters in granite bedrock, may shed further light on Gold's hypothesis.

In searching for new fields, drillers seeking gas and oil in traditional suspect source reservoirs, during the 1960s and 1970s, would find gas and/or oil in only about 9 of 100 wells drilled and usually only 2 or 3 of these wells produced sufficient gas and/or oil to be of commercial value. Whereas the average depth of gas well drilled during this period ranged between 5000 and 6000 feet (1524–1829 meters), some drillers are now aiming at the 30,000-foot (9144-meter) depth. For many years, the deepest gas well in Texas was 28,600 feet (8717 meters). In the early days of offshore drilling, operations were conducted in water only 20–25 feet (6–7.5 meters) deep. there are now many platforms in waters that are deeper by a factor of 20–30 times.

In terms of the conventional or traditional sources, natural gas is found in areas close to exposed or buried mountain ranges. Major deposits of natural gas are found in inclined strata where the rock formations dip away from the crest of a buried hill or the ridge of a buried mountain. Some of the common types of formations in which natural gas is found are shown in Fig. 1. Although natural gas and crude oil are frequently found together, the largest natural gas reserves (about 70% of the estimated reserves) are in deposits neither in contact with, nor dissolved in, oil.

Forecasting Natural Gas Reserves

Since the 1920s, natural gas reserves have been based upon the amount of gas that most likely will be found with oil. Although numerous experts in the field have claimed over the years that this methodology overlooks large amounts of "unassociated" gas, the professionals have been slow to revise their procedures. The concept that much natu-

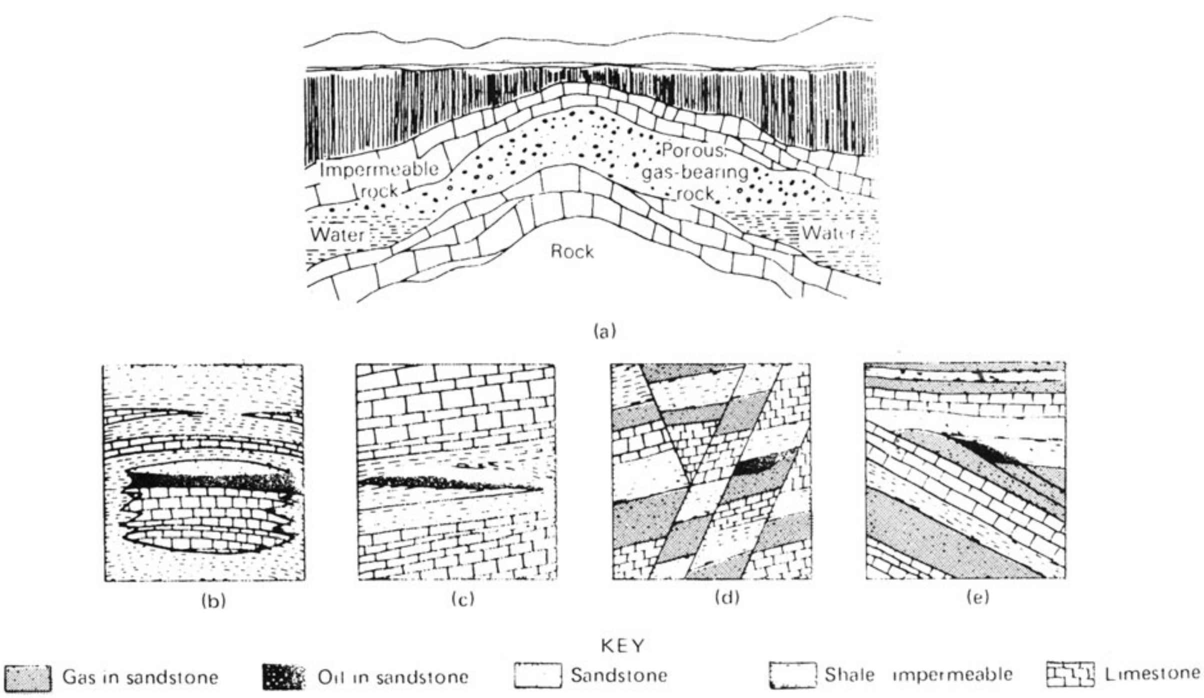


Fig. 1. Types of natural gas reservoirs and entrapments: (a) anticlinal trap; (b) coral reef trap; (c) stratigraphic trap; (d) fault trap; and (e) unconformity.

ral gas occurs quite apart from oil is now taking hold. The U.S. Geological Survey, which has proclaimed a gas resource base at about 400 trillion cubic feet, is now undertaking a study to be completed in about 1996. Survey techniques have undergone several dramatic improvements, including the use of computer techniques that will project three-dimensional survey information, as contrasted with past reliance on two-dimensional mapping. A number of experts now believe that the reserves are well over 1,200 trillion cubic feet (34 trillion cubic meters). One independent gas producer has estimated the figure at about 1,500 trillion cubic feet (42.5 trillion cubic meters).

Traditional estimates of oil-associated natural gas reserves historically have rated the former Soviet Bloc as holding nearly 40% of the reserves, Iran about 14%, and the United States nearly 6%, with additional fairly high reserves in Qatar, Algeria, and Saudi Arabia. At one time, North America was attributed to have a 60-year supply, but with revisions in estimating procedures, coupled with increased efficiency in natural gas production and gas combustion efficiency, the future is now believed to be in terms of at least a few centuries. Interest continues, however, in upgrading low-Btu natural gases and developing so-called *substitute natural gas*.

The principal gas fields within the continental United States are indicated in Fig. 2.

Fig. 2. Location of principal natural gas fields in the lower 48 of the United States. Alaska, not shown, ranks third in terms of holding estimated reserves. (*Batelle Memorial Institute*.)

Ultimate Recovery of Natural Gas Reserves. It has been traditional for many years to categorize ultimate recovery of gas reserves by reservoir lithology which involves the compilation of such data by three types of reservoirs: (1) *Sandstone reservoir*—consisting of sedimentary rock composed predominantly of quartz grains or other noncarbonate mineral or rock detritus. Included in this reservoir type are unconsolidated sand, sandstone, siltstone, graywacke, arkose and granite wash, conglomerate, and breccia. (2) *Carbonate reservoir*—composed of sedimentary rock made up predominately of calcite (limestone) and/or dolomite. (3) *Other reservoirs*—including igneous and metamorphic rocks and some sedimentary rocks, such as fractured shale.

Estimated ultimate recovery is also reported by type of entrapment, of which there are two major types: (1) *Structural trap*—an entrapment in which migration of hydrocarbons in the reservoir rock has terminated primarily because of closure induced by structural deformation, such as folding or faulting. Within this category should also be included entrapments attributed to hydrodynamic forces. (2) *Stratigraphic trap*—an entrapment in which migration of hydrocarbons has terminated because of the pinchout of reservoir rock due either to truncation or to nondeposition or to a facies change in the form of diminished permeability of reservoir rock. Also included in this category are entrapments in which a pinchout of facies change provides part of the barrier to migration of hydrocarbons, with structural elements providing the remaining closure for the entrapment. In these cases, it is recognized that the dominant cause of the accumulation is the lenticularity of truncation of the reservoir rock.

Some estimates indicate that about 65.8% of ultimately recoverable natural gas in the United States will be found in structural traps; the remaining 34.2% in stratigraphic traps.

The estimated ultimate recovery is also reported by the geologic age of the reservoir. It is recognized that problems may arise where the geologic age of a reservoir cannot be determined specifically, such as Permo-Pennsylvanian and Cambro-Ordovician, or where production from reservoirs of different geologic age are combined.

Natural gas liquids occur in either the gaseous phase or in solution with crude oil in the reservoir. They are recovered at the surface as liquids by separation from produced natural gas, by such processes as condensation and absorption in field separators, gasoline plants, and other surface facilities. In this processing, valuable by-products are recovered, such as light oils, natural gasoline, and other petroleum gases such as ethane, propane, and butane. Natural gasoline is blended with gasoline from petroleum refineries to improve starting properties, especially desirable in cold weather. Ethane is a major petrochemical raw material. Propane and butane are made available as LPG (liquefied petroleum gas). Processing of natural gas also removes unwanted material, such as nitrogen, sulfur compounds, carbon dioxide, and water vapor. Some gas fields produce helium, which is extracted cryogenically.

Unconventional Sources of Natural Gas. These include: (1) tight sandstones, (2) Devonian shales, (3) geopressured zones, (4) deep basins, (5) gas associated with coal seams, and (6) gas in the form of methane hydrates.

1. *Tight Sandstones.* In the United States, tight sandstones of the western basins range from the northern tier states to the Mexican border. Some tight gas sands also occur in the eastern United States. To date, resource development has occurred only in the limited areas characterized by thick, fairly uniform, blanket-type formations which, when hydraulically fractured, provide sufficient gas production rates to merit commercial exploitation. In these areas, as pointed out by Sharer (Gas Research Institute, GRI), state-of-the-art technologies can be used because only a limited knowledge of the formation characteristics is required for economic production. However, a majority of the resource base is associated with lower permeability and more complex blanket and lenticular sand formations, for which current technology is not adequate. GRI is concentrating research in these areas.
2. *Devonian Shales.* The large eastern Devonian gas shales resource base underlies approximately 174,000 square miles (453,000 km^2) of the eastern U.S. Estimates of recoverable gas range from 2 to 15% of the gas in place. Natural gas has been produced from these shales for decades. Well production rates are relatively low, but after the first few years of production it does not usually decline rapidly with time. A major constraint to present-day exploitation has been the extraordinary inability to predict with confidence the gas production rates that may be obtained in wells drilled outside the traditional production areas. Presently, the GRI is studying the systematics of historically successful fields, including the Appalachian, Illinois, and Michigan Basins.
3. *Geopressured Zones.* A test well in a geopressure zone was drilled some years ago in Tigre Lagoon in the coastal marshes of southern Louisiana. Known as Edna Delcambre #1, this well produced at a rate of up to 10,000 barrels of water per day from a sandstone aquifer some 12,600 feet (3840 meters) below the surface. Pressure at that depth is nearly 11,000 pounds per square inch (748 atmospheres) and the temperature is 116°C. Quite an elaborate manifold system is required to collect the gas. The water is disposed by forcing it by its own pressure into another well bore which penetrates to a depth of 2500 feet (762 meters). Scientists associated with this project had expected about 20 cubic feet/barrel (42 gallon); about 0.6 cubic meter/barrel (159 liters). In actuality, reports indicate that the yield of gas was about 2.5 times that amount.

As explained by specialists in geopressure technology, at great depths (in terms of present technology), the solubility of natural gas in water may be as much as 1000 cubic feet/barrel (28.3 cubic meters/159 liters) at depths of 30,000 feet (9144 meters), whereas that solubility will be reduced by a factor of ten at a depth of 20,000 feet (6096 meters). Under the right combination of geologic and hydrologic conditions, this gas-laden water will move toward the surface, during which process some of the gas will be released from the water in the form of very small bubbles. Ultimately, this gas collects beneath a geologic trap, where conventional free-gas reservoirs are formed. Some authorities now believe that the very deep aquifers are much more extensive than the free-gas reservoirs. It is this gas-saturated water that some scientists believe will be a great source of future natural gas.

The GRI has been investigating the coproduction of gas and water for a number of years. Natural gas from watered-out reservoirs, geopressured aquifers, and high-water-saturated gas-bearing reservoir strata are prime targets. This natural gas is trapped by water such that special production techniques must be used to move the water and remobilize the gas. Although some gas is also dissolved in the water, it is of less significance than the free gas trapped as dispersed bubbles or in pockets or stringers of various sizes in the reservoir rock matrix.

4. *Deep Basins.* These are found at depths between 15,000 and 30,000 feet (4572–9144 meters) and are estimated to contain significant quantities of gas, but generally await the development of advanced production technology and economic incentive.
5. *Gas Associated with Coal Seams.* Methane, the principal constituent of natural gas, is generated during the geologic process of coal formation. A significant portion of this gas is trapped by impermeable strata and present within the fractures and micropores of the coal. (The presence of methane is an ever-present hazard in coal mining.) Major variations in resource estimates are due to uncertainties in the gas content and size of the deeper, unminable coal deposits in the western states that form the major portion of the resource base. Seeking such gas may involve depths as great as 6000 feet (1829 meters) underground. Except for reasons of safety, little effort has been made to recover any of this resource due to high recovery costs, potential uncertainties in production, and deficiencies in state-of-the-art equipment, particularly for the deeper coals. The GRI is concentrating its research efforts on unminable coal because of its large potential as a resource base. While the gas resource associated with mining amounts to about 10% of the energy value of the coal, producers rarely apply new gas recovery technology except where safety is a requirement. Targets of the GRI program are deep coal seams, multiple seams interbedded with shales and sandstones, and deep multiple beds that are too thin to mine.
6. *Methane Hydrates.* Within a certain range of pressures and temperatures, methane and water form hydrates. Described as icelike substances, these hydrates are believed to occur in very substantial quantities, particularly beneath permafrost and in deep-ocean bottoms. Although slush has occurred in gas pipelines under certain conditions for many years, the existence of hydrates in nature was not made known until the mid-1960s. Geologists and hydrologists had previously assumed that gas of this type would have dissipated during earlier geologic ages. This is another area of natural gas resource research awaiting economic incentives.

Exploratory Methods. The principal exploratory methods used are: (1) *Airborne magnetometers*, which seek out anomalies in the magnetic field. Experienced geologists relate these irregularities to the probability of gas reservoirs below the surface. See also **Magnetometer.** (2) *Satellite imagery*, from which surface structures and patterns can be related to previous pattern recognition studies made of surfaces below which gas reservoirs exist. (3) *Gravitometers* are used to detect subtle variations in gravitational pull inasmuch as this is less for a gas reservoir than for continuous dense rock formations. (4) *Seismic methods* which constitute the most widely used of exploration methods. See also **Earth Tectonics and Earthquakes.** (5) *Data logging methods*, wherein an instrument is lowered into the borehole and which telemeters back to the surface readings of sonic absorption in an effort to determine the nature and thickness of rock formations. Data loggers operate on the basis of several physical phenomena. (6) *Fossil inspection*. The careful examination of microfossils can assist in fixing the age of rocks that are being penetrated. The condition of the fossils also can be related to probable temperatures to which they have been exposed over geologic periods and these, in turn, can be advantageous in locating possible gas deposits. Usually a combination of two or more exploratory techniques is used.

Liquefied Natural Gas (LNG)

The liquefaction of natural gas for storage and transportation and regasification for final distribution dates back several decades. A few major accidents in the handling of LNG thwarted the progress of the field for a while, but in the early 1970s, LNG was again considered in a major way because of energy-short nations. One of the more serious LNG accidents occurred in Cleveland, Ohio on October 20, 1944, when a storage tank developed a leak with spillage and subsequent fires in the surrounding neighborhood in which 135 persons lost their lives. While liquefaction offers marked storage space savings and convenience, the predominant advantage occurs in connection with both pipeline and ship transportation. Energy-short nations, such as Japan, and some of the European nations, have turned in recent years to the concept of shipping LNG by ship. For example, a large LNG plant at Lumut, Brunei, Borneo went onstream in mid-1974 essentially to furnish LNG to Japan.

Oil- and gas-rich nations, which at one time flared to the atmosphere much of the natural gas that accompanied the production of crude oil, have turned toward conservation—either through reinjection of much of the natural gas underground or through constructing LNG production facilities for shipment of the product overseas. Concurrent with such planning was reevaluation by a number of nations of their own valuable resources and a growing reluctance toward exporting inordinate quantities of gas and oil strictly for money. As of the early 1990s, the shipment of LNG overseas competes with other ways and means for alleviating energy shortages, including coal conversion and gasification, nuclear energy, solar energy, etc.

Three types of liquefaction processes may be used for production of LNG. The standard cascade process which uses three refrigerants—methane, ethylene, and propane—all circulating in closed cycles, is shown in Fig. 3. There is a separate compressor for each of these re-

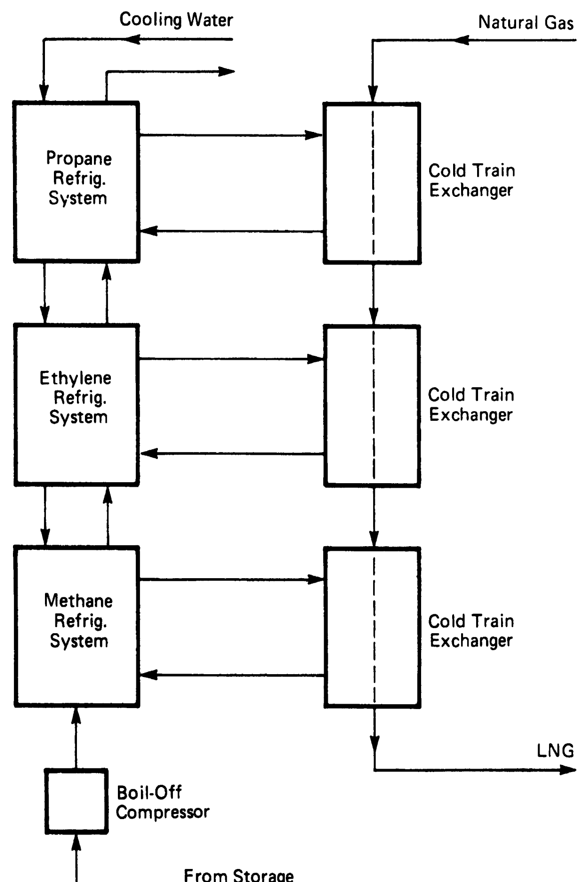


Fig. 3. Conventional or standard cascade system for producing liquefied natural gas (LNG).

frigerants. The methane and propane are available from the feed gas (natural gas). The ethylene must be furnished separately. Ethane may be used in place of ethylene at a subatmospheric suction pressure. The cascade process has the highest rank in terms of thermal efficiency. As a possible improvement over the cascade process, the mixed refrigerant process was developed in the early 1960s. A single-pressure mixed refrigerant cascade (MRC) system is shown in Fig. 4. In one plant using this process, a hydrocarbon-plus-nitrogen mixture of relatively wide boiling range (N_2 through C_5) is used as the refrigerant. All of these components can be recovered from natural gas in separate apparatus. In still another system, shown in Fig. 5 a propane and mixture-refrigerant cycle is used. In this process, the cooling load is divided horizontally at about −34.4°C into an upper portion absorbed by propane and a lower portion absorbed by the mixed refrigerant. In essence, the system is a dual refrigerant cascade in which the lower boiling fluid is a mixture refrigerant. The cascade combination with propane makes it possible to reduce the boiling range of the mixture refrigerant substantially, which improves the thermodynamic efficiency over that of the straight MRC process.

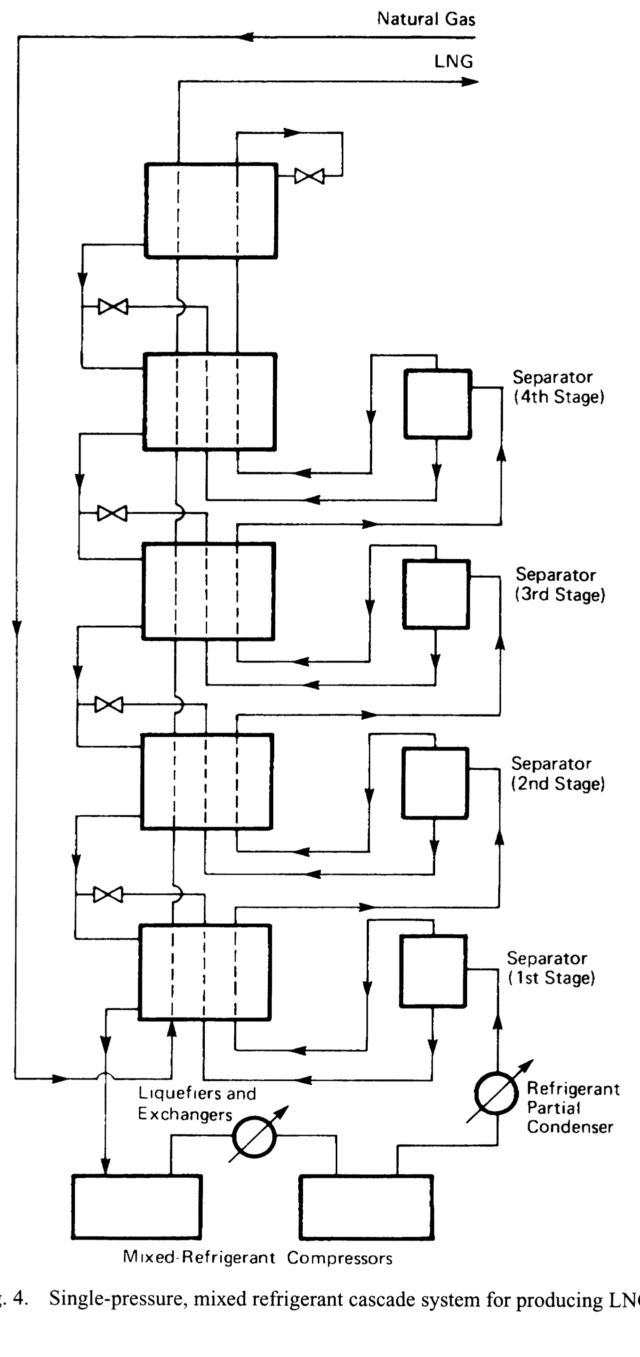


Fig. 4. Single-pressure, mixed refrigerant cascade system for producing LNG.

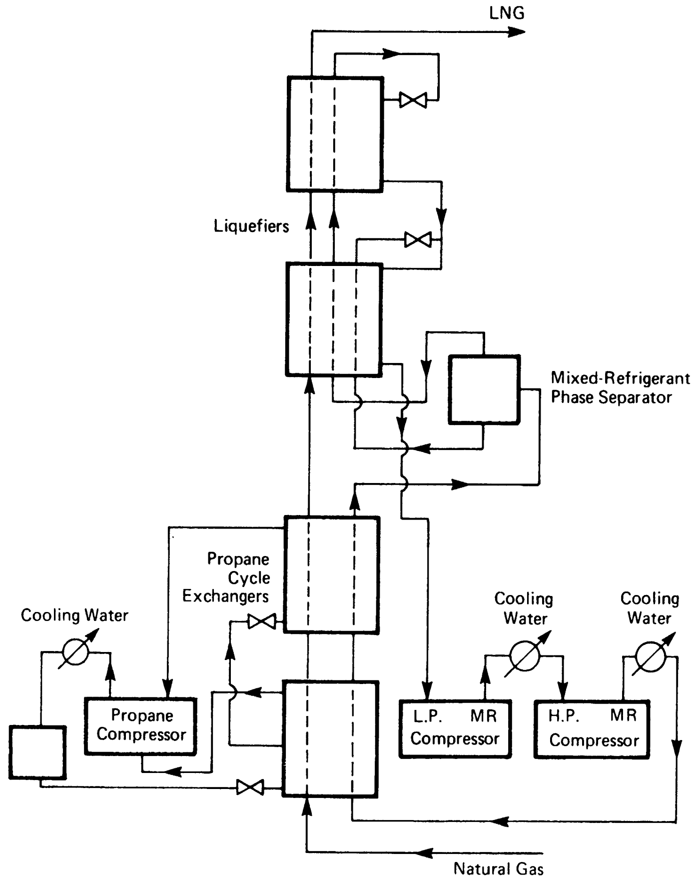


Fig. 5. Propane-mixed-refrigerant liquefaction system for producing LNG. L.P. = low pressure; H.P. = high pressure; MR = mixed refrigerant.

Cryogenic Upgrading of Low-Btu Natural Gases

Worldwide, there are substantial reserves of natural gas in which the reservoir formation hydrocarbons are contaminated with nonburning components. The presence of components, such as helium, nitrogen, or carbon dioxide, reduces the heating value of the gas mixture and can result in the gas being unsuitable for existing transmission and distribution systems, Such contaminated mixtures are termed low-Btu gases if their heating values fall below the minimum standards, regulations, or contract heating value requirements.

Cryogenic processing can be used to upgrade some of these low-Btu gases so as to produce an acceptable high-Btu product. Cryogenic upgrading is a physical process in which subambient temperatures are employed to bring about a separation between the hydrocarbons and nonhydrocarbons in the mixture. The reduction of temperature occurring during cryogenic processing produces a two phase (gas-liquid) mixture. The relative volatilities between the components in the mixture result in selective mass transfer between the two phases. One phase becomes enriched with hydrocarbons and then has a heating value higher than the original gas. The second phase becomes denuded of hydrocarbons and has a heating value below that of the original gas mixture. Frequently, the mass transfer operation requires several theoretical stages in order to achieve the desired product heating value and high hydrocarbon recoveries. While cryogenic upgrading can be applied to gas mixtures containing carbon dioxide or hydrogen sulfide, it has so far only been applied commercially to those hydrocarbon mixtures contaminated with nitrogen and helium.

One of the main considerations in the design and operation of cryogenic upgrading plants is to identify and remove any component from the gas which could adversely affect the operation of the cold sections of the plant. Such components are carbon dioxide, water vapor, and heavy hydrocarbons which have high solidification temperatures and low solubilities. In general, if these components are allowed to remain in the gas to the cryogenic unit, they will form solids during the cooling process which will be deposited on the heat exchanger surfaces. This will lead to fall-off in performance and, possibly, to blockages and plant shutdown.

Particular attention should be paid to identifying any high freezing point components in the low-Btu gas. There exists a range of absorption and adsorption processes to pretreat the low-Btu gas to remove these undesirable components.

A simplified flowsheet of cryogenic upgrading plant is given in Fig. 6. The plant, consisting of two identical trains, is capable of processing 260 million standard cubic feet (7.3 million cubic meters) per day of low-Btu gas (580 Btu/standard cubic foot) (5162 Calories/cubic meter) and upgrading the gas into 143 million standard cubic feet (4 million cubic meters) of high-Btu gas (980 Btu/standard cubic foot; 8722 Calories/cubic meter). The plant stream parameters are indicated in Table 2.

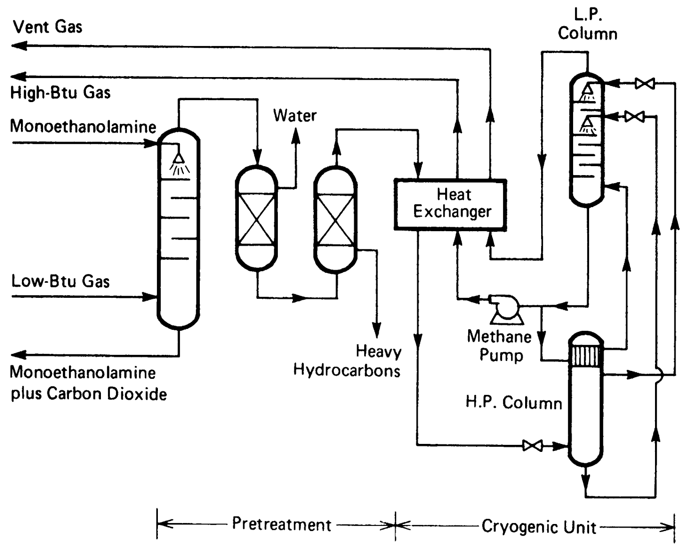


Fig. 6. Plant for nitrogen removal from natural gas using cryogenic upgrading. (*Petrocarbon Developments, Ltd.*)

TABLE 2. CRYOGENIC UPGRADING OF NATURAL GAS— STREAM PARAMETERS Composition—Mol.%

	Low-Btu Gas	High-Btu Gas	Vent Gas	Helium
Helium	0.40	—	0.09	100.00
Nitrogen	42.75	4.00	98.95	—
Methane	56.02	95.09	0.96	—
Ethane +	0.53	0.91	—	—
CO_2	0.30	—	—	—
Flow (million standard cubic feet/day)	246	143	100	0.43
Flow (million cubic meters/day)	7	4	2.8	0.012
Heating value (Btu/standard cubic foot)	580	980	—	—
Heating value (Calories/cubic meter)	5162	8722	—	—

The low-Btu gas is available at 800 psig (54 atmospheres) and is mainly a nitrogen-methane mixture. In addition to a small quantity of helium, the gas also contains small quantities of carbon dioxide, water vapor, and heavy hydrocarbons. The carbon dioxide is removed by washing with monoethanolamine (MEA), the water is taken out on molecular sieve, and the heavy hydrocarbons by adsorption on activated carbon. The gas is then cooled in aluminum plate-fin exchangers against the returning high-Btu product gas and vent gas. The gas is then expanded to 380 psig (26 atmospheres) and a vapor-liquid mixture passes into the H.P. (high pressure) fractionator. The purpose of this fractionator is to bring about an initial separation of the nitrogen-methane and to produce a liquid reflux for the L.P. (low pressure) fractionator.

A nitrogen-enriched vapor flows up the H.P. fractionator, while methane is returned to the sump of this column by a nitrogen reflux stream produced in the tubes of the overhead condenser. The refrigeration required to produce this nitrogen reflux is provided by evaporating some of the liquid methane, from the L.P. column, in the shell of the overhead condenser. Two liquid streams are taken from the H.P. fractionator, and these become the feed and reflux for the L.P. fractionator. The L.P. feed is an enriched methane stream taken from the base of the H.P. fractionator. The L.P. reflux is a high purity nitrogen liquid taken off the H.P. fractionator just below the condenser. The upgrading is completed in the L.P. fractionator. The feed stream is stripped to produce a high-Btu liquid containing 4% nitrogen and having a heating value of 980 Btu/standard cubic foot (8722 Calories per cubic meter). The liquid is pumped from the column sump, evaporated, and superheated against the incoming low-Btu gas. The gas from the top of the L.P. column is mainly nitrogen and is also heated to ambient temperature against the incoming low-Btu gas. By using this arrangement of two distillation columns, the separation of nitrogen and methane can be achieved using only the pressure energy available in the low-Btu gas.

Transportation of Natural Gas

The mode selected for gas transportation depends mainly on (a) the distance over which the gas must be moved; (b) the geographical and geological characteristics of the terrain (considering both overland and overseas [underseas]) across which the gas must be moved; (c) environmental factors directly associated with the gas transportation mode; and (d) the physical characteristics of the gas to be transported, notably, the phase—whether gaseous or liquid; and (e) the construction and projected operating costs of the transportation system, based upon trading off the advantages and limitations over which some flexibility of selection may be present. Aside from economic factors, a system can be engineered to transport either the gaseous or liquid phase, thus giving rise to considerable flexibility in certain situations.

Overland Pipelines. Detailed maps of gas pipelines in the United States and other parts of the world can be found in several references, particularly among the periodicals which serve the pipeline industry. Notable among these references is the international petroleum encyclopedia and atlas issued periodically by Petroleum Publishing Co., Tulsa, Oklahoma. Numerous trade associations serving the pipeline industry are also excellent sources on pipeline statistics. There are so many pipelines that presentation of this type of information is beyond the scope of this encyclopedia.

Historically, Texas, Louisiana, Oklahoma, and New Mexico have been large producers of natural gas, as well as some significant fields in the West Virginia-Ohio-Pennsylvania area. New developments in Alaska are and will continue to influence the gas transportation and distribution pattern.

Much of the installed gas pipeline ranges from 14 to 30 inches (36 to 76 centimeters) in diameter, the most common ranging from 20 to 35 inches (51 to 89 centimeters) but there is a strong trend toward larger-diameter lines, from 42 inches (107 centimeters) upward. Line pipe is made from high-strength plates, $\frac{3}{8}$ inch to 1 inch (1 to 2.5 centimeters) in thickness. Sections of pipe are usually 40 feet (12 meters) long, minimum, ranging up to 60 or 80 feet (18 or 24 meters). Lengths of pipe arrive at the scene most often by truck and are strung out by special pipe carriers along the right-of-way so that the construction crews will find them near the place where they are to be installed. Helicopter delivery of pipe is sometimes used where it is impossible for trucks to do the job. The total weight of steel going into a long-distance pipeline is impressive. For example, a pipe with a wall thickness of $\frac{1}{2}$ inch (13 millimeters) and a diameter of 30 inches (76 centimeters) will weigh more than 400 tons (360 metric tons) per mile.

In building very long pipelines, the pipeline company usually employs several construction contractors. The total length of line is divided into a number of sections with separate equipment and crews. Usually, each crew works on not more than 100 miles (161 kilometers).

By partitioning the construction task, the entire operation can be speeded up, particularly important in areas where freezing temperatures or rain and mud may interfere with the work.

The numerous machines needed to dig the trench, weld the sections of pipe, apply protective coating to prevent corrosion, lower the pipe into the trench, and cover the trench with earth are known collectively as a *main line spread*. The trench is usually 3 or more feet (1 meter or more) in depth, sufficiently deep to prevent damage by plowing and earth-moving equipment. Depending on the size of the pipe, the trench will range from 2 to 4 feet (0.6 to 1.2 meters) or more in width.

Teams of welders join the pipe sections into a continuous tube. The most modern welding techniques involve automatic welding machines. X-ray equipment is used to inspect welds. When several sections of pipe have been welded together, the continuous tube is lowered gently into the trench by *sideboom tractors*. These machines have cranes or derricks slanted over to one side so that they can pick up the pipe and lower it several feet away from the tractor itself. Pipe purchased from steel mills may come with a coating and wrapping already applied. The thick coating may be of coal tar or asphaltic material, which is then covered with heavy paper or fiberglass. This protective coat-and-wrap is needed to prevent rusting. If bare pipe is used, there are special machines that coat and wrap right on the job just before the pipe is lowered into the trench.

A special piece of equipment, known as the *holiday detector*, is a hoop of metal which is placed around the pipe after it is coated and wrapped. A small electrical current flows through the hoop. If there is a "holiday," i.e., a spot where there is no coating, the detector alerts the operator. This is brought to the attention of a special crew that coats and wraps bare spots in the pipe.

Since pipelines do not follow an absolutely straight line, bending machines are used to curve the pipe in the vertical, horizontal, or both directions. When a pipeline must cross a river, the contractor will dig or dredge a deep trench in the river bed. The pipe is then surrounded by heavy weights and encased in concrete so that it will not be carried away by the current. If there is a suitable bridge across the river, the pipeline may be hung from the underside of the steel girders of the bridge. In some cases, a special bridge is constructed to carry the pipe across the stream. In crossing a highway or railroad, the pipeline must be put through a tunnel under the structure. A giant auger will be used to bore under the road to accommodate a section of somewhat larger-diameter pipe, forming the tunnel through which the main pipeline passes.

Gas pressures in long-distance pipelines may range from 500 to 5,000 pounds per square inch (34 to 340 atmospheres) with 1,000 psi (68 atmospheres) being quite common. Pressure is boosted to make up for frictional losses by use of compressor stations located every 50 to 100 miles (80 to 161 kilometers) along the pipeline. In terms of lineal velocity, natural gas may travel at a rate of about 15 miles (24 kilometers) per hour; thus, about three days are required to move a molecule of gas over a distance of 1,000 miles (1609 kilometers).

All along the pipeline, there are valves and regulators that may be opened or shut to control the internal pressure, or to cut off the flow entirely if an unexpected break in the line is caused by a flood, earthquake, or other disaster. The valves and regulators can be operated by microwave radio long before any crew could reach them. Stations for reducing the pressure, located near points of consumption, frequently are called *city gates*. These stations measure the amount of gas leaving the main pipeline at this point as well as reducing the pressure.

Marine Pipelines for Gas

With some alterations, the techniques that apply to construction and laying of marine pipelines for gas also apply to fluids, such as oil. Marine pipelines can be underwater in a river, marsh, or ocean, but the predominant industry effort in recent years is the construction of pipelines in the open ocean at increasingly deeper levels. The trend toward deepwater pipelining and construction in harsher environments naturally follows the expansion of the search for offshore gas and oil. This search began in earnest after World War II and is expanding at an ever-increasing rate; even if slowed to some extent by some environmental concerns in the United States, the rate is rapid in other parts of the world. Worldwide energy needs have caused oil companies to move into areas that only a few years ago would have been too expensive to develop on a practical basis. Lines are now being laid in water depths of several hundred feet (meters) and cover distances of 200 miles (320 kilometers) or more from field to shore. These longer lines are major trunklines bringing gas and oil to land terminals. Other lines are necessary out at the field to connect platforms to each other; or possibly to connect platforms to sea berths.

The sizing of the pipeline, the design of the pumping and compression systems needed to move the products, the design of the automation systems, and many of the corrosion control procedures are the same regardless of whether the pipeline is on land or at sea. The two major areas of design difference between land and marine pipelines are (1) the stresses incurred in getting the pipeline to the sea bottom, and (2) the necessity of keeping the line stable and in place while it is exposed to forces induced by current and wave.

The stability problem is theoretically simple but is complicated somewhat by the uncertainty of precise values for some of the coefficients used in the calculations. Basically, it is a matter of providing enough weight in the pipe and pipe coating system to provide a net downward force when balanced against the buoyance and the lift force caused by the seawater moving by the pipe. This net downward force, in conjunction with the coefficient or friction for the particular pipe-soil combination under examination, can then mobilize a horizontal resisting force. This should be somewhat larger than the drag force exerted on the pipe by the water motion in order to give the desired safety factor.

Different safety factors or horizontal water velocities may be utilized depending on the operating conditions that will be encountered during the life of the pipeline. For example, many pipelines will be buried beneath the sea bottom at some time interval, ranging from a few weeks to a year or two, after their construction. The exposure of this line to maximum horizontal water velocities caused by storm current and waves is obviously much less than that of a line that will remain on the surface of the sea bottom. It is also obviously necessary to consider whether the line will contain gas, oil, or other substance at the time the design loads may occur.

It is important to carefully consider the foregoing points in the design of the weight coating since the ability of a contractor to safely construct the line relates very closely to the negative buoyancy of the pipe and coating.

The most common method of marine pipeline construction utilizes a floating vessel on which the pipe is assembled in a horizontal position. As additional joints or sections of pipe are added to the already-completed segment, the barge is moved forward, actually moving out from under the completed pipeline. This is sometimes called the "stovepipe" method, named after the manner the pipe sections are added, one after another. This pipeline extends off the stern of the vessel and spans down to the sea bottom. It is supported part of the way down by a construction aid called a "pontoon" or "stinger." This is basically a slender structure pinned to the vessel on one end and with built-in buoyancy that can be controlled so that it floats at the proper angle to the water surface to provide support to the pipeline.

In shallow water, the pipeline is then allowed to span from the end of the pontoon to the sea bottom as a simple beam. As water depths increase, it becomes necessary to add tension to the pipe on the barge. This, of course, changes the analytical problem from one of a simple beam to one of a beam under tension. This analysis must take into account the weight of the pipe, the wall thickness, the type of steel in the pipe, the tension on the pipe, the support of the pontoon, the geometrical configuration of the tension on the pipe, the geometrical configuration of the pipe-pontoon-barge system, and the pipe end condition at the sea bottom.

There are three basic configurations of pipelay vessels in common usage: (1) the barge-type hull; (2) the ship-shape hull; and (3) the semi-submersible vessel. The barge-type hull (Fig. 7) is the most common because of its economy and simplicity, its ability to provide the space and stability for heavy lifts and deck cargo, including pipe, and its shallow draft, permitting work close inshore. The primary disadvantage is its relative sensitivity to sea conditions. In particular, roll and heave motions will shut down pipelay operations in 6-foot to 14-foot (1.8- to 4.2-meter) waves, depending on wave direction and period.

Overseas Shipping of LNG. A key feature of most LNG carriers in operation is the insulation system which maintains the cargo at −

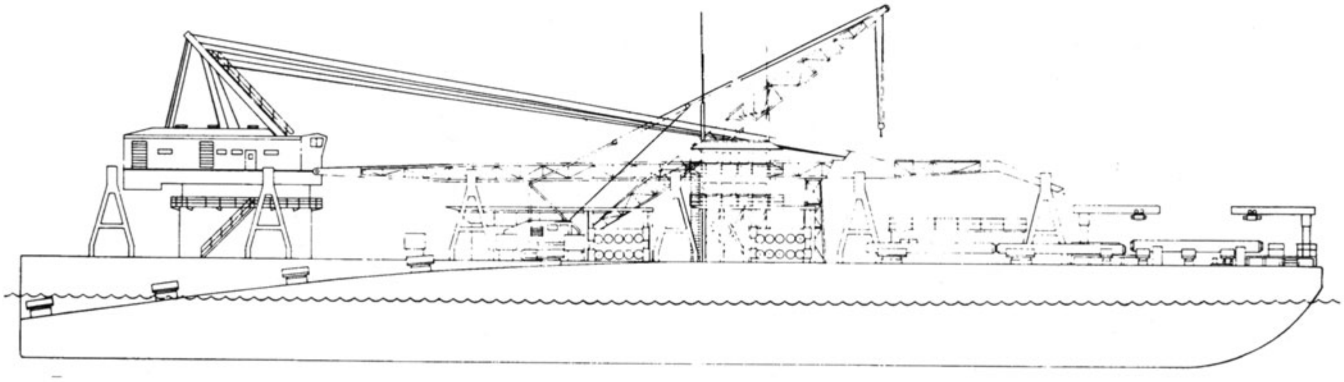


Fig. 7. Pipelay vessel with barge-type hull.

162°C. In one type of ship, the cargo is carried in five tanks constructed of a thin welded membrane of special steel. Each tank is separated from the inner hull by insulating material. The small fraction of the cargo that boils off because of heat leakage is used as boiler fuel for the propulsion of the ship. On a loaded voyage, this may provide about 90% of the fuel needed. The ships are ballasted for return voyage with seawater carried in separate wing tanks. Some LNG is left in the cargo tanks to ensure a nonexplosive gaseous atmosphere and to keep the tanks cool for the next voyage. Again, boil-off gas provides part of the propulsion fuel. One configuration of an LNG ship-loading system is shown in Fig. 8.

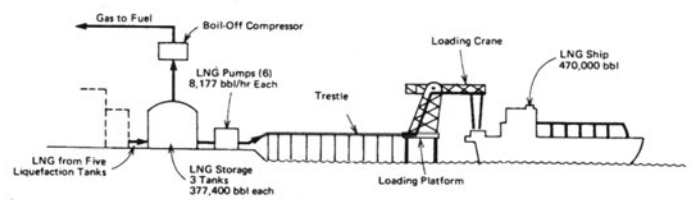


Fig. 8. LNG ship-loading system. Trestle is 2.6 miles (4.3 km) long.

Safety in handling LNG in ships and at loading and unloading terminals of large scale has been a matter of constant concern. The observation has been made that the LNG gas carried would bury a football field under 125 feet (38 meters) of liquefied gas, or, after conversion to the gaseous phase, 600 football fields to the same depth. One factor that has not been routinely considered in the past is a phenomenon called a *flameless vapor explosion*. It is well known that if water, for example, could be heated without nucleation occurring on the sides of the vessel, the water temperature could be raised well above the boiling point of 100°C. If this could be done, and with the continued application of heat, the liquid would suddenly explode in its transition from the liquid to the vapor phase. Although not probable, it is possible that conditions favoring flameless vapor explosion could occur if liquefied natural gas were permitted to escape over a water surface. One scientist has observed that an explosion of this kind is possible when a liquid is 4 to 6% (no more, no less) above its normal temperature of vaporization. It is further observed that an explosion of this nature would not occur when LNG first spreads across a volume of water, but with time and the warming of the LNG such a hazard could occur. While an explosion of this type is not comparable to that from a chemical reaction, the explosion could greatly disperse the LNG over a greater area, thus spreading the zone of risk.

Underground Storage

The largest additional supply of natural gas for peak demands comes from underground storage reservoirs located, for example, close to the northern cities, as compared with the producing wells which may be located in the southwestern area of the country. Some of the storage pools are operated by pipeline companies, but most of the gas in underground storage is owned by the local gas companies that serve metropolitan areas.

The underground reservoirs are filled with gas from the pipelines during the summer months, when all of the fuel that the lines can deliver is not consumed. This method allows the producing wells and the pipelines to operate at fairly steady rates at all times of the year. Also, it is established that a gas field will produce more gas over a longer period if the gas is withdrawn at a steady rate.

Four states—Michigan, Pennsylvania, Illinois, and Ohio—have half of the total underground gas storage pools in the United States, with a total capacity of over 5 trillion cubic feet (142 billion cubic meters). In a typical year, about one-fourth of this volume will be used during cold waves to furnish the additional gas needed to supply homes and apartments.

The most common type of underground reservoir now storing gas is a previously producing gas or oil field. The supplies remaining in these pools are too small, and at too low a pressure, to justify continued production. But, the reservoir rock can hold gas pumped down through the same wells that once took gas out of the ground.

About 90% of the storage pools being used once produced gas or oil. In Pennsylvania there are over 60 such pools close to the large industries and centers of population. There are over 30 such pools in West Virginia, Michigan, Ohio, Kansas, Indiana, New York, and Kentucky, as well as smaller numbers in 13 other states. The gas to be stored is pumped into the old wells by compressors similar to those used to move gas in pipelines. The gas is stored under about the same pressure as originally existed in the field. In developing a gas storage reservoir, a company obtains a lease from the landowners in much the same manner that gas producers do.

The gas industry has been developing underground storage reservoirs for more than 60 years. The first known experiment in storing gas underground was conducted in 1915 in Welland County, Ontario, Canada by the National Fuel Gas Company. The success of this effort prompted the Iroquois Gas Corporation, a subsidiary of National Fuel, to develop, in 1916, the Zoar field south of Buffalo, New York. It was the first storage operation in the United States and is the oldest continuously used reservoir.

During the past 60 years, over 80 companies have invested several billions of dollars in underground storage facilities.

Another kind of underground storage reservoir is called an aquifer. An aquifer is an underground rock structure holding large quantities of water. The underground rock is porous and permeable. The pore spaces are filled with water, and impermeable rock covers the porous rock. Wells are drilled into such formations, and gas is forced into the pores under pressure. As the gas pressure increases, the gas pushes the water farther down into the porous rock, making room for the gas.

There are over 40 aquifers in the United States, located in Illinois, Indiana, Iowa, Kentucky, Minnesota, Missouri, Utah, and Washington. Three unusual reservoirs have been developed: an abandoned coal mine in Colorado; and salt domes in Michigan and Mississippi.

History of Natural Gas as an Energy Resource

It is reported that, perhaps 2,000 years ago, the Chinese piped natural gas from shallow wells through bamboo poles, for burning under large pans to evaporate seawater for salt. The first commercial use of natural gas in the western world was for lighting the streets of Genoa, Italy, circa 1802. The first evidence of natural gas deposits in the United States is found in reports of "burning springs" in various parts of New York, Pennsylvania, Ohio, and West Virginia. As early as 1626, French missionaries visiting the Indians in northwestern New York recorded that they could ignite gases rising from shallow waters. Many early reports were given of the presence of natural gas along the shores of Lake Erie and in the streams flowing into it. There also are references to "burning springs" in the Ohio River valley and along the Pacific shores of California. It is reported that General George Washington was fascinated by a "burning spring" in the Kanawha Valley, near Charleston, West Virginia, in 1775. Early settlers who drilled wells for water often reported the presence of traces of natural gas. The generally accepted birthplace of the natural gas industry in the United States is Fredonia, New York. Fredonia is located on Canadaway Creek, which empties into Lake Erie in the northwest corner of New York State. William A. Hart is reported to have dug a well in 1825 and obtained sufficient natural gas to light two stores, two shops, and a grist mill. Hollow logs were used for piping. Sufficient gas would accumulate in the well riser during the day to supply the gas lights at dusk. Hart was also instrumental in building the first natural gas lighthouse in 1829 along Lake Erie. The lighthouse, consisting of 13 gas lamps and reflectors in two tiers, served until 1859. In 1858, the first natural gas company in the United States was formed, The Fredonia Gas Light Company.

The consumption of natural gas gradually increased prior to World Wars I and II as more and more small pipelines brought communities within reach of natural gas fields accompanied by the retirement of previous manufactured or town gas facilities (the early forerunners of the substitute natural gas).

Natural Gas–powered Vehicles. The concept of natural gas–powered vehicles has become a *limited* reality in terms of the millions of gasoline- and diesel-fluid-power vehicles. As of 1993, Mack Trucks (Allentown, Pennsylvania) and the Gas Research Institute have teamed to research and develop a natural-gas version of the Mack E7™ heavy-duty engine. A prototype of the design is scheduled for testing on a refuse vehicle in the Boston area. If successful, the engine also could be applied to a variety of heavy-duty vehicles, including long-haul tractor/trailers, construction equipment, and road maintenance trucks. The development is propelled by the needs of truck fleet owners who may be required by legislation to operate alternatively fueled vehicles. The engine will be required to meet applicable U.S. Environmental Protection Agency and California Air Resources Board emissions standards while maintaining the performance and reliability of its diesel-fueled counterpart. A 6-cylinder, 12-liter engine will be developed. The vehicle's onboard gas storage will hold the energy equivalent of about 45 gallons (170 liters) of diesel fuel.

Substitute Natural Gas. The oil crisis of the 1970s spawned a number of attempts to create synthetic natural gas. Several of these processes are described in the Sixth Edition of this Encyclopedia. Some of these processes reached pilot and demonstration plant stages and beyond in their development. For example, substitute natural gas (SNG) from sewage wastes has enjoyed impressive success. The anearobic digestion of a solid waste and water or sewage sludge slurry will produce a methane-rich gas.

Additional Reading

Abelson, P. H.: "The Gas Research Institute," *Science*, 1715 (December 11, 1992).

Bethke, C. M., et al.: "Supercomputer Analysis of Sedimentary Basins," *Science*, 261 (January 15, 1988).

Burnett, W. M., and S. D. Ban: "Changing Prospects for Natural Gas in the United States," *Science*, 305 (April 21, 1989).

Fischetti, M.: "There's Gas in Them Thar Hills!" *Technology Review (MIT)*, 17 (January 1993).

Fulkerson, W., Judkins, R. R., and M. K. Sanghvi: "Energy from Fossil Fuels," *Sci. Amer.*, 136 (September 1990).

Holtberg, P.: "1993 Policy Implications of the GRI Baseline Projection of U.S. Energy Supply and Demand to 2010," Gas Research Institute (Washington Operations), Washington, D.C. (1993).

Jensen, B. A.: "Improve Control of Cryogenic Gas Plants," *Hydrocarbon Processing*, 109 (May 1991).

McCabe, K. A., Rassenti, S. J., and V. L. Smith: "Natural Gas Pipeline Networks," *Science*, 534 (October 25, 1991).

Melvin, A.: "Natural Gas: Basic Science and Technology," Adam Hilger (London), Taylor & Francis (Philadelphia), 1988.

Woods, T. J.: "The Long-Term Trends in U.S. Gas Supply and Prices: 1992 Edition of the GRI Baseline Projection of U.S. Energy Supply and Demand to 2010," Gas Research Institute, Chicago, Illinois, December 1991.

NATURAL RUBBER. See **Rubber (Natural).**

NAUTICAL MILE. The fundamental unit of distance used in navigation, once defined, for purposes of convenience, as 6,080 feet (1,853 meters). Rigorously, the nautical mile was defined as the length of 1 minute of arc on a great circle drawn on the surface of a sphere with the same area as the earth; thus as 6,080.27 feet (1,853.27 meters).

But, owing to the fact that the earth is an oblate spheroid and flattened at the poles, the length of 1 minute of arc measured along a meridian varies in different latitudes. It is shortest at the poles and longest at the equator, having an average length of 6,076.82 feet (1,852.21 meters). To resolve the confusion caused by this variation of the nautical mile with latitude, various countries established standard figures. At last an international agreement was reached whereby the nautical mile was defined, and is presently accepted, as 1,852 meters. For navigation purposes, a good approximation for the number of nautical miles is the number of minutes of arc along a great-circle route.

The nautical mile is frequently confused with the *geographical mile*, which is defined as the length of 1 minute of arc on the earth's equator and has a length of 6,087.15 feet (1,855.36 meters). See also **Units and Standards.**

NAVIER-STOKES EQUATIONS. The equations of motion of a Newtonian fluid which are applicable to the motion of simple liquids and non-dissociating gases. In tensor form, they are

$$\frac{du_i}{dt} = \frac{\partial u_i}{\partial t} + u_j \frac{\partial u_i}{\partial u_j} = -\frac{\partial p}{\rho\, \partial x_i} + \mu \frac{\partial^2 u_i}{\partial x_j^2} + \frac{\mu}{3} \frac{\partial}{\partial x_i}\left(\frac{\partial u_i}{\partial x_i}\right)$$

where du_i/dt is the time rate of change of the velocity of a fluid with velocity vector u_i, u_j is the velocity vector perpendicular to u_j, ρ is fluid density, p is the hydrostatic pressure, and μ is viscosity. See also **Fluid Flow.**

NAVIGATION. Since development of the early navigation satellites (INMARSAT, et al.) just a few decades ago, the approach to the science and equipment of navigation has undergone a major anatomical change that is geared to ultrasimplification for the end user. The operator no longer requires sophisticated support instrumentation or an understanding of and appreciation for the geometry and mathematics that have typified pre-satellite navigation systems. Refinements of navigation satellite systems, as exemplified by the Global Positioning System (GPS), first applied by the military during the Persian Gulf War (1990–1991), are occurring apace and ultimately to a significant extent will obsolete many of the early and contemporary navigation systems. Because of the installed investment in contemporary systems, possible security restrictions that may be imposed on GPS during wartime, and the ability of some users throughout the world operating on limited budgets, the time of phaseout for contemporary non-satellite–based systems is unknown. Further, there is much interest in the lore of navigation methodologies. This article incorporates, as in prior editions, a chronology of navigation system developments. The article commences, however, with a condensed overview of the GPS system.

Space-Age Triangulation

In the Global Positioning System, satellites with precisely synchronized atomic clocks continuously broadcast their time and location to suitable earthbound computerized receivers. In a fully self-contained manner, the receiver calculates and displays its Earth position. Data received from at least three satellites is required to yield the *latitude* and *longitude* of the receiver. Data from a fourth satellite is required to yield *altitude*. The receiver is synchronized to the satellites' clocks. Distance from the receiver to the satellite is calculated by measuring the time interval required for the signal to travel from satellite to receiver. When three satellites are in "view" simultaneously, a computer in the receiver quickly works outs the precise position by triangulation because only one point on a plane can yield a particular combination of distances from the three satellites. See Fig. 1. Receivers are small and may be handheld.

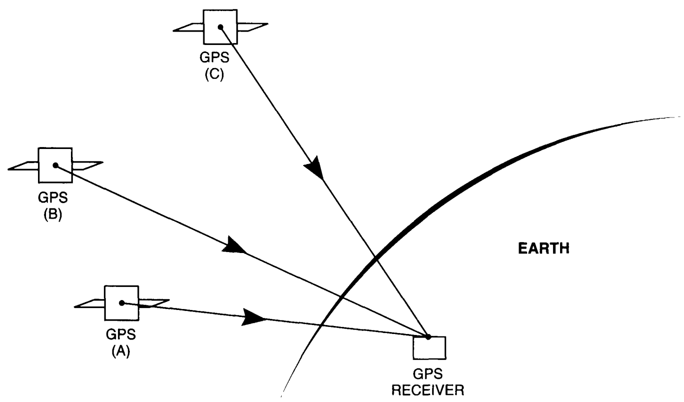


Fig. 1. Global Positioning Satellites (A, B, C), each equipped with a precisely synchronized atomic clock, continuously broadcast their time and location to earthbound receivers. Computers in the receiver calculate distances of receiver from each satellite, this information derived from the time required for transmission of signal from satellite to receiver. By employing the methodology of *triangulation,* the receiver determines the only *acceptable location* that will geometrically satisfy the signal time differences. Only three GPS readings are required to determine *latitude* and *longitude*. Additional data from a fourth GPS will yield *altitude* information. The GPS thus combines advanced satellite technology with simple geometry.

Quite a large number of satellites in orbit are needed to ensure that at least three satellites are in sight at any particular Earthpoint and four in view if altitude information is required as well. As of 1994, the U.S. military intends to have 24 satellites in place, including three spares. Each satellite orbits Earth once every 12 hours at an altitude of 10,900 miles (17,538 km). The ultimate total cost of the GPS system is estimated at $10 billion upward.

Because of the advantages of the GPS to air, sea, and ground operations during the military venture Desert Storm, equipment was released for use in the field at least 2 years prior to an established schedule. The GPS markedly increased the accuracy of bombing runs, accurately guided sea-launched missiles to targets, steered ships and personnel safely through mine fields, and advised tank and land-based commanders with reliable information on their exact location even in featureless desert scenarios.

Had the enemy been aware of more detail receiver design and had developed suitable receivers, they could have taken advantage of the GPS. Subsequent analyses of the various encounters did not reveal, however, any enemy use of the system. But this possibility in terms of future military actions has prompted the U.S. military to offer GPS information on two levels by way of an arrangement known as *differential GPS.* For most military uses of GPS, the maximum available accuracy is needed. Reports on early design of the system indicated that accuracy could be expressed in terms of a few millimeters rather than meters. An early satellite-based system had provided scientists in various fields, notably in connection with movement of Earth's crust, with a *fine signal* in the millimeter range.

Although subject to future alteration, the military developed the concept of *selective availability*—that is, (1) a fine and very accurate position reading, and (2) a course location fix, the accuracy of which would be in the range of 100 meters. In this system, a slight distortion of the satellite signal is introduced but still permits a finer, undistorted signal for specially designed receivers for military applications. This approach basically denies location information of the best achievable accuracy to scientists and for a host of demanding applications of the system for commercial use. One scientist has observed, "The military has come up with a system that is useful to mankind. Now they don't want mankind to have full use of it." This scientist is engaged in studies for predicting seismic risks.

For example, commercial uses that require information of high accuracy are found in the monitoring of railway cars on tracks that are quite close to each other and for navigating ocean-going vessels through rock-infested straits where, at a minimum, an accuracy of a few meters is required.

Some scientists have developed electronic means to circumvent the signal distortion problem by using schemes that are identified as *differential GPS.* In one version of this technique, GPS information from numerous satellites and using several receivers will cancel out minor errors. One scientist has observed, "We've learned to live with selective availability, but the solution is costly, by perhaps an order of magnitude increase in the GPS readings required. This in turn is reflected by increases in the costs of data storage, transmission, and analysis."

Although GPS is a major breakthrough for navigation and position-related analyses, the system, like communications satellites, can be plagued with satellite instability problems, faulty antenna pointing, and failure of solar panels for power, among other factors.

The early work required to develop navigation satellites, dating back to the 1970s and ultimately leading to the GPS, is described in considerable detail in the Seventh Edition of this encyclopedia.

Classification of Navigation Sensors

The term *navigation* stems from the Latin word *navigare,* meaning "to conduct a ship." Traditionally, navigation has been treated under three main classifications: (1) the *sailings*—to find the direction or directions in which to head a ship in order that it may proceed on the most practical course from one place to another (see also **Sailings (The)**), (2) *dead reckoning*—to find the position of a ship in any particular instant, provided that its position at some prior time is given and that, since leaving this position, the headings, distances run on each heading, and the effects of environment and motions of the medium supporting the ship are known (see also **Dead Reckoning**), and (3) *celestial navigation*—to find the position of a ship at any instant by means of observations, not necessarily visual, of one or more objects on the surface of the earth, or on the celestial sphere. Although still a helpful classification, the foregoing obviously is dominated by consideration of ships at sea, not taking into full account air and space navigation. Also, within the last few decades, navigational instrumentation has been put onto a much more sophisticated footing so that a classification of the type shown in Fig. 1 is much more useful.

Further, there are two broad categories of navigation: (1) *absolute navigation*, wherein knowledge of present position is known in relation to an overall earth-coordinate system (latitude and longitude, for example); and (2) *relative navigation*, wherein position is known relative to some special local coordinate or grid system. The accuracy attainable in a relative navigation situation usually is significantly better than that which can be obtained on an absolute basis.

Navigation Sensors

The primary navigation sensors and their functions are listed in Table 1. The more commonly used sensors for both marine and aircraft navigation are included. Frequently, a navigation system will be made up of various combinations of these sensors.

As will be noted from Fig. 2, there are self-contained navigation systems (self-processed information) and externally-controlled systems wherein a data link is required between the vehicle (aircraft, ships, etc.) being navigated and some reference or data point. An abridged list of examples of these two general classes of systems is given in Table 2.

Many factors are involved in selecting the most appropriate navigation system, including the usual parameters of accuracy, cost, size, weight, power requirements, reliability, operational simplicity/complexity, degree of automaticity desired, ruggedness in environmental extremes—factors which have given rise to numerous standards of equipment. Other very important factors for many situations include, for example, worldwide availability, all-weather operability, and continuous versus periodic availability. Certain radio navigation systems, for example, may not embrace transmitters that cover the entire world. All-weather, night-and-day operation rules out, for example, such sys-

TABLE 1. NAVIGATION SENSORS BY PRIMARY FUNCTION

Meters

Velocity Meters	Altitude or Depth Meters
Pilot log	Pressure altimeter
Electromagnetic log	Pressure-depth meter
RPM tachometer	Radar altimeter
Airspeed (pilot) meter plus wind	Sonar fathometer
Doppler radar	Inertial systems
Laser radar	Ground or other vehicle
Doppler sonar	Trackers and data link
Inertial systems	

Time Meters
Chronometer
Radio-synchronous signals
Time standards (crystal oscillators)

Heading Reference

Magnetic compass	Radiometric celestial tracker
Gyro compass	Inertial systems
Optical celestial tracker	

Position-fix Devices

Sextant	Navigation satellite	Radio transponders
Radio aids	Landmarks	Ground or "mother" ship
LORAN	Optical sighting	Trackers and data link
OMEGA	Radar sighting	Gravity anomaly, map matching
SHORAN	Seamarks	Magnetic anomaly, map matching
RAYDIST	Sonar bottom fixing	
DECCA	Optical celestial tracker	
VOR/DME TACAN	Radiometric celestial tracker	

tems as optical star trackers (as sole sensors); clouds below an aircraft, for example, rule out laser doppler systems. In some navigational satellite systems, there may not always be a satellite within range, and a fix may have to await radio view when one comes into position. Thus, it is rare when a single navigation sensor can satisfy all desired requirements. Even magnetic compasses and normal-mode gyrocompasses are essentially unusable in the polar regions of the Earth.

Celestial Navigation

In this system, navigation is achieved by means of observing celestial objects. In the year 1837, Captain Thomas Sumner discovered what has since been known as the Sumner Line, and modern celestial navigation may be said to date from that discovery. The methods for determining terrestrial latitude and longitude from sextant observations of altitude of celestial objects are briefly described in the entry on **Sumner Line.** During intervening years, much research has been carried out on the theory of celestial line of position, and numerous methods for calculating the data necessary to plot that line have been developed.

It is in order to briefly describe what is meant by a celestial line of position. At any instant, any celestial object is directly at the zenith for some particular spot on the surface of the earth. This point in years past was known as the subsolar, sublunar, or subastral point, depending upon whether the object observed was the sun, the moon, or a planet or star. The term *Ground Position* (GP) is now used no matter what celestial object is used. The terrestrial coordinates of GP may be expressed in terms of celestial coordinates of the object, the latitude being equal to the declination of the object, and the longitude equal to its Greenwich hour angle. Both of these quantities are tabulated for the sun, moon, planets, and navigator's stars in various readily available almanacs. The tabulations are given in terms of Greenwich Civil Time, and this time is used by navigators for recording the times of sextant observations for altitude.

In practice, the line of position is drawn on a Mercator chart, a plotting sheet, or a small-area plotting sheet by employing the geometric proposition that a radius of a circle is always perpendicular to an arc. At some particular Greenwich Civil Time (GCT), the altitude (h_s) of a celestial object is obtained with the sextant or bubble octant. Then, using the dead-reckoning position, or some position close to it, which leads to simplified computations, the values of the altitude (h_c) and the bearing (Z_n) that the object would have at the assumed position and GCT of observation are computed or taken from suitable tables. The dead-reckoning, or assumed, position is now set down on the plotting sheet and a line drawn through it in the direction Z_n. This line is a section of the radius of a circle drawn about the GP, which, in reality, is usually off the plotting sheet. Next, the difference between the computed and observed altitudes ($h_c - h_t$) is taken; this is called the "intercept." If the intercept is zero, the ship must be on a line of position perpendicular to the bearing line and passing through the plotted position. If the intercept is plus (+), the line of position must pass through a point ($h_c - h_t$) minutes of arc, or nautical miles, away from the GP along the bearing line; and if the intercept is minus (−), the line must be between the plotted position and the GP. In either case, the line of position must be perpendicular to the bearing line.

In spite of the fact that the line of position is actually a circle with radius ($90° - h_t$) miles, the line may be drawn as straight in practically all cases. If we assume that the altitude of the object is 80°, the value of ($90° - h_t$) is 10° or 600 miles (965.6 kilometers). In this case, a straight line 60 miles long (96.56 kilometers), perpendicular to the radius, will differ from the actual circle by less than a mile at its extremities. Accordingly, the assumption of a straight line will not lead to appreciable errors, if the altitude is less than 80° and the drawn line is less than 60 miles (96.56 kilometers) long.

Simple statistical analysis shows that the point on the line of position closest to the dead-reckoning point is the most probable position that can be obtained for the ship from a single observation of altitude. In air navigation, this position is referred to as the estimated position (EP). Care must be taken not to confuse this EP with the EP obtained by dead-reckoning navigation in marine navigation. This most probable position, or EP, can be obtained without plotting the line by any of the above methods if the dead-reckoning (DR) position instead of the assumed position is used, since the intercept gives the shortest distance from the DR position to the line.

An example of the use of celestial navigation at sea is given in the following practical case:

During the night, the navigating officer of a ship on passage from England to the United States wishes to check the dead-reckoning position of the ship. The two stars Alpheratz and Altair are well placed for observation. When the navigating officer's watch reads 23h 40m 10.0s, the sextant altitude of Alpheratz is 50° 34′.3. For the purpose of checking the deviation of the steering compass, the bearing of the star is taken by this compass and found to be 121°. At watch reading 23h 46m 15.4s, the sextant altitude of Altair is 50° 20′.7. The watch times must be corrected to obtain Greenwich Civil Time (GCT), and the sextant altitudes corrected for instrumental errors, dip, and refraction to obtain the geocentric altitude (h_t). These corrected results are:

Star	GCT	h_t
Alpheratz	02h 39m 34.0s	52° 27′.4
Altair	02h 45m 39.4s	50° 13′.8

Using the average of the watch times, the dead-reckoning position of the ship is found as latitude 43° 24′.6N and longitude 48° 27′.4W. Since the ship is proceeding at only 16 knots, and since the DR position is probably somewhat in error, this position is used for computing the altitudes and bearings that the stars should have at the GCT observation. These values are found to be:

Star	Bearing	h_c
Alpheratz	098°.4	52° 32′.5
Altair	216°.2	50° 07′.4

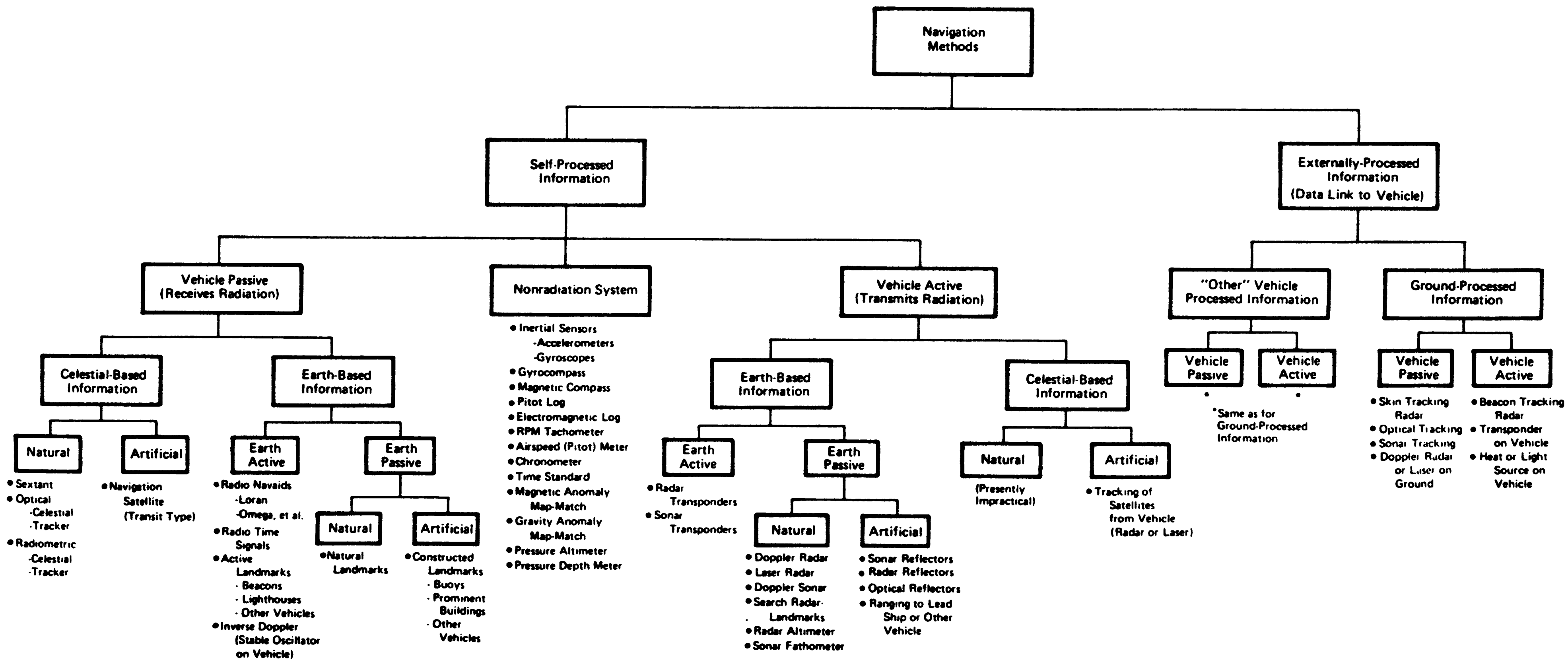


Fig. 2. Classification of sensors used in navigation systems.

TABLE 2. SELF-CONTAINED AND EXTERNALLY-CONTROLLED SYSTEMS

SELF-CONTAINED SYSTEMS
—Completely self-contained
—Self-processed information; no radiation required.
  All measurements internal or in local vicinity of vehicle, such as inertial navigation.
—Self-processed information
—Vehicle is passive; natural earth or natural sky references.
  Natural can include constructed landmarks (buildings, etc.) not purposely placed as navigation aids.
  Optical star trackers.

SELF-CONTROLLED SYSTEMS
—Self-processed information; vehicle passive; earth passive or artificial
—Purposeful navigation landmarks.
  Navigational buoys.
—Self-processed information; vehicle passive; active earth or active sky
  LORAN.
—Self-processed information; vehicle active; earth passive or artificial; or artificial celestial
  Ranging to lead ship.
—Self-processed information; vehicle active; earth active
  Radar transponders.

EXTERNALLY-CONTROLLED SYSTEMS
—Ground or other vehicle processed information and data link; vehicle passive; external tracker passive
  Theodolite optical tracker.
  Skin tracking radar.
—Ground or other vehicle processed information and data link; vehicle active; external tracker passive
  Optical tracking of light on vehicle.
—Ground or other vehicle processed information and data link; vehicle active; external tracker active
  Beacon tracking radar.

Foregoing are examples only; not all-inclusive.

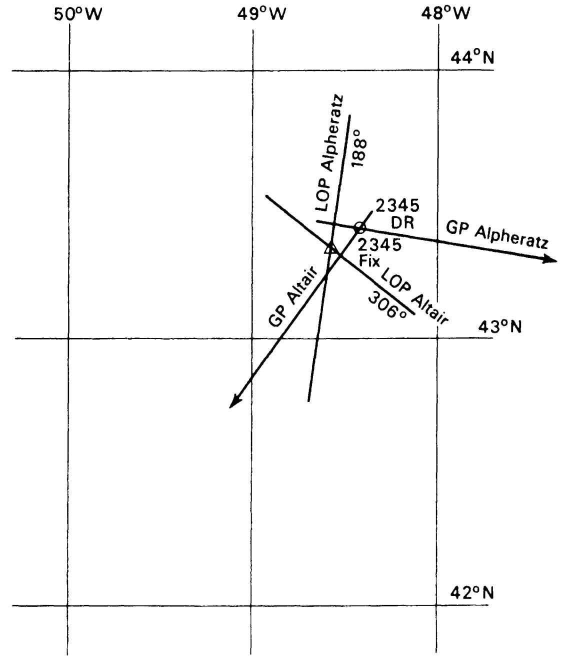


Fig. 3. Scale drawing for celestial navigation example.

The intercepts ($h_c - h_t$) are found: for Alpheratz $+5'.1$ and Altair $-6'.4$. These yield two "most probable" positions of the ship, one 5.1 miles from the DR position in the direction 278°.4 (i.e., away from the GP of Alpheratz), and the other 6.4 miles in the direction 216°.2 (i.e., toward the GP of Altair). To determine the fix, the DR position is plotted, the bearing lines to the two GP's drawn through it, the intercepts measured off in the proper direction, and the two lines of positions drawn through the intercepts perpendicular to the lines to the GP points. The point of intersection of the lines of position is the fix at 2345. The actual plotting, on small area plotting sheet, is shown in Fig. 3. From the figure, the fix is found to be in latitude 43° 20'.7N and longitude 48° 35'.2W. To determine the compass deviation, the difference between the observed compass bearing of Alpheratz and the computed bearing is found to be 121° − 098° = 23°. Since the variation in this region is 26°W, the deviation must be 3°E.

Proper selection of stars to be observed will yield data of extreme importance to the pilot and navigator. For example, if the object is nearly ahead or astern of the plane, the line of position will cross the course nearly at right angles, and the length of the intercept will provide a check on the ground speed being made good. On the other hand, if the object is in a direction approximately perpendicular to the course, the value of the intercept will indicate the accuracy of the wind correction angle.

In many cases, altitudes and bearings of celestial objects may be computed in advance of the actual observing. These predetermined altitudes have many uses in air navigation. If an aircraft is to depart at a definite time and follow a specified course, the altitudes and bearings at indicated times may be computed before the plane leaves the ground. The course to be followed is plotted on a chart, the predetermined DR positions for indicated times are marked, and the bearings of the GP of the object are indicated by lines drawn through the DR positions. The precomputed altitudes are geocentric, but they may be transformed into those expected to be read to the octant at the specified times by applying the various corrections with reversed signs. The navigator measures the altitude at an indicated time, obtains the difference between his value and that predicted, and lays off this distance along the drawn bearing line either toward or away from the GP. In this way, an EP is determined within a few seconds after the observation is completed, and no computing is required during the flight. The pilot is notified to alter heading and air speed to bring the plane back to schedule. If, due to unforeseen conditions, the plane gets so far off scheduled position that the intercepts are more than 150 miles, the predetermined altitudes must be abandoned and regular celestial navigation adopted.

Under some conditions, it may be necessary for a plane to make an accurate landfall (e.g., locate a small island, life raft, etc.) under conditions where celestial navigation must be relied upon. In such cases, the use of precomputed altitudes gives great assistance. First, an estimated time of arrival (ETA) is obtained. Then, using the latitude and longitude of the landfall, a series of altitudes and bearings of a celestial object is computed. The interval of time between computed values depends somewhat on the rapidity with which the values are changing, but is usually about 10 minutes. The series begins at least half an hour before ETA and extends beyond that value. Two curves are then drawn on graph paper, showing altitude and bearing as a function of time. The plotted altitudes are those expected with the octant, i.e., with corrections applied to computed geocentric values. If possible, an object that is approximately ahead or astern of the plane should be selected. About half an hour before the predicted ETA, the pilot alters heading 10° or 15° to the right or left of that predicted for the true course, so that there will be no question as to which side of the landfall he is approaching, and the navigator begins taking altitudes of the object. The navigator plots his observed values, as a function of time of observation, on the same graph as that showing the predetermined values, and obtains a curve of observed values. At the instant that the observed curve intersects the curve obtained from precomputation, the plane must be on a line of position running through the landfall. The bearing of the celestial object at this instant is read off the plotted bearing curve, the line is drawn at right angles to this bearing through the destination, and the pilot is instructed to alter heading to run down the line.

Radio Navigation

The use of radio aids in navigation for checking the dead-reckoning position of a ship is known as radio navigation. Radio direction-finders were used very early in the development of radio technology to avoid the difficulties of celestial navigation from a ship or aircraft and for emergencies in bad weather. The simplest system uses a directional antenna to locate the direction to several radio stations. A simple triangulation then locates the ship or aircraft with respect to the location of the stations, usually well known and in the map of reference with which the pilot is familiar. The method is complicated by the aircraft velocity, but not by accelerations, weather, or the availability of tables. The pilot can tune in a station near his destination and simply follow the signal to it. Two deficiencies are present: (1) the location of the stations; and (2) the errors inherent in a directional antenna. These problems led to improved radio systems.

The use of radio bearings as lines of positions is best explained by an example from ship navigation. A ship is proceeding on heading 330° at 12 knots. Three radio beacons, *A*, *B*, and *C*, are located in the following positions:

Station	*A*	*B*	*C*
Latitude	29° 30′N	30° 00′N	28° 40′N
Longitude	83° 20′W	81° 40′W	81° 52′W

At 0812 the dead-reckoning position of the ship is L = 28° 32′N & Lo = 82° 42′W and at that time, radio bearings, corrected for deviation of the radio compass, are A = 000°, B = 063°, and C = 117°. These must be changed to true bearings by adding to each the heading of the ship, obtaining: A = 330°, B = 033°, and C = 087°. Since these are great-circle bearing, they must be reduced to rhumb-line bearings by applying the correction factors for $A - 0.05$, $B + 0°.2$, and $C + 0°.1$. Then, working either on a mercator chart, mercator plotting sheet, or small-area plotting sheet, the corrected rhumb-line bearings are plotted, and the fix determined as the center of the triangle of intersection of the three lines of position. The position of the fix is L = 28° 37′N & Lo = 82° 44′W, and, since the sides of the triangle are less than 3 miles, we can assume that the fix is probably correct to within 1 mile. The complete solution is illustrated in Fig. 4 which is drawn on a small-area plotting sheet and labeled in accordance with standard procedure.

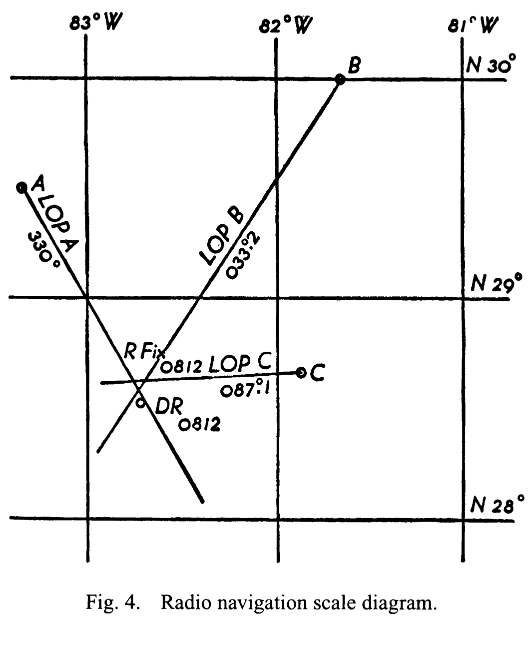


Fig. 4. Radio navigation scale diagram.

Three corrections must be applied to a radio bearing before it can be used as a line of position on a mercator chart or small-area plotting sheet. Loop antennae and radio direction-finders have deviation corrections, due to the magnetic field of the ship. These must be determined in advance for different headings of the ship and applied to radio bearings as obtained. Then the radio bearing, which is relative, must be changed to true bearing by adding the true heading of the ship. Finally, since radio follows great circles, the bearing must be converted from great-circle to mercator, or rhumb-line, bearing. Figure 5 shows two

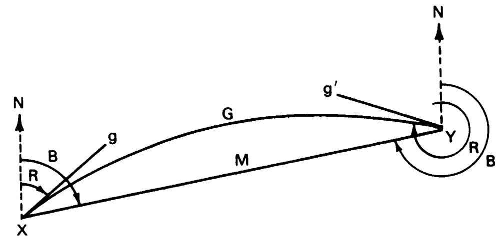


Fig. 5. Plotting of *X* and *Y* on a mercator chart.

points, *X* and *Y*, plotted on a mercator chart, with the rhumb line, *XMY*, and the great circle, *XGY*, connecting the two points. The great circle will always be convex toward the nearest pole, and we have drawn the figure for the Northern Hemisphere with true north indicated both at *X* and *Y*. The lines *gX* and *g*′Y are tangents to the great circle at *X* and *Y*, respectively, and are the directions in which the signal from *Y* will arrive at *X*, and that from *X* will arrive at *Y*. Let us consider *X* to be the receiving station. Then the angle *R* (*NXg*) represents the great-circle bearing of *Y*, and the angle *B* (*NXM*) the rhumb-line bearing. In this case, it is noted that a correction must be added to the great-circle bearing to obtain the rhumb line. Reversing stations and considering *Y* the receiver, we see that at this point the correction must be subtracted to obtain rhumb line from great circle.

In case a navigator is working on a Lambert chart, as is frequently the case in air navigation, the great-circle bearing is close enough to a Lambert line to be plotted without correction other than for deviation and heading.

A radio range is a system of radio signals designed for the purpose of guiding a ship or plane along a designated track toward or away from a specified location. Relatively low frequency (200–400 kHz) has been used. Although there are various modifications, the track is indicated by the intersection of two field patterns from the range antenna system. The usual antenna arrangement is two pairs of cross antennae set 90° in space from one another. This gives two figure-eight field patterns as shown in Fig. 6 (a). The patterns overlap in narrow wedge-shaped regions which have their apices at the transmitting station. These overlapping sectors are known as the range, or "the beam." Both an aural and visible system may be used for keeping a ship "on the beam."

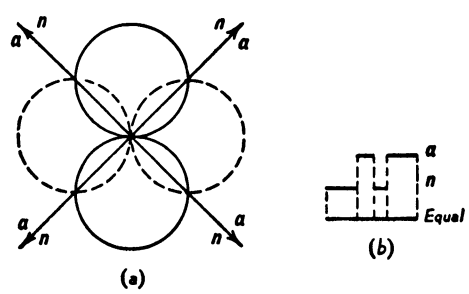


Fig. 6. Radio ranging.

In the aural system the carriers from the two antennae are so modulated with some audio note, say 1,000 Hz. The code signal (letter *a*, dot dash) is transmitted from one antenna, and the code signal (letter *n*, dash dot) from the other, so timed that on the center of the overlapping region the two signals blend together in a continuous note. This is indicated in part (a) of Fig. 6. and the time sequence of the code characters is shown in part (b). To remain directly "on the beam," all that is necessary is for the navigator to head his plane so that the continuous note is heard in his receiver. If he drifts to the right or left of the center, he hears either the *a* or *n* signal superimposed on the steady note. This also provides a method for planes proceeding in opposite directions to remain on the proper side of the airway. The range signals are interrupted at frequent intervals to give the name of the station and, when necessary, weather reports and information to aviators or navigators.

In the visual method, the carriers from the two antennae are modulated with different frequencies, say 65 and 85 Hz. The antennae are then excited alternately so that the plane receiver gets first one signal and then the other very rapidly. The demodulated output of the receiver is fed to a tuned reed instrument so if the two signals are received with equal strengths, indicating on course, both reeds will vibrate with equal amplitude, while off-course flight will cause a greater vibration of one or the other depending upon which side of the course the plane is flying The exact angular relation of the courses laid down by the range station may be altered in several ways. Feeding the two antennae with different strength signals, feeding in different phases, utilization of additional antenna elements, etc., all serve to alter the field pattern so the lines of equal strength can be varied in direction. Where it is desired to rotate the courses after their angular relation has been fixed, a double goniometer may be used to feed the antennae. This gives a continuous 360° control of the direction of the beams.

The term *radio compass* has been used loosely over the years. When the loop antenna was first applied to the determination of radio bearings, the term *radio-compass station* was applied to shore installations that would forward, upon request, the bearing of a ship from the station. Next, the term was applied to a group of shore installations, each equipped with a loop antenna, from which the navigating officer of a ship within range could obtain the latitude and longitude of his ship. After the loop antenna and receiving sets had been developed to a state where they could be carried by the ships themselves, the term radio compass was applied to the loop. As new and improved radio equipment became available, the term radio compass was successively applied to any radio device that could be used to determine bearing. A glance through any textbook on navigation, particularly those dealing with air navigation, will yield at least two, and sometimes as many as five, different instruments for radio compass.

Radio compass is also applied to a direction-finding instrument which has a dial and looks in many respects like an ordinary compass and which is used for heading the ship in much the same manner as a compass. Two radio antennae are used with the instrument—a loop and a nondirectional antenna. The volume controls for signal intensity are so adjusted that the signal strength from the loop and the nondirectional antenna are the same when the station is in the plane of the loop. The two antennae are then fed into a single receiver, and the signal intensity is illustrated in Fig. 7. The resultant signal intensity from any station is shown on the instrument panel.

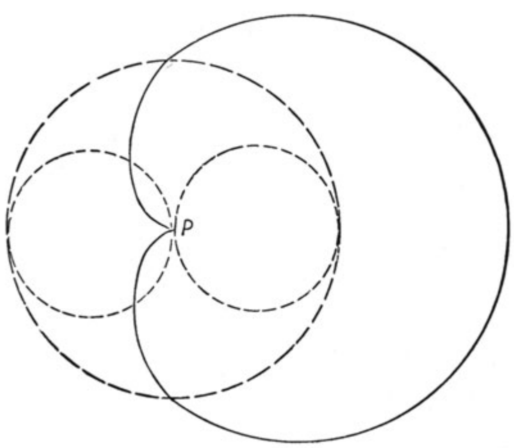


Fig. 7. Radio compass antenna patterns.

With reference to the figure, the length of a line from P to any one of the three curves is proportional to the strength of the signal from a station in the direction toward which the line is drawn. The dotted (figure eight) curve represents the relative intensities in various directions for the loop alone; the dashed (circle) curve represents the intensity for the nondirectional antenna alone; and the full curve that for the combined loop and nondirectional antenna, i.e., for the radio compass.

Hyperbolic Navigation. Hyperbolic navigation is a general method for determining a line of position by measuring the difference in the distance from the navigator of two stations of known position. The difference in distance is found by measuring the difference in time between arrival of signals transmitted from the two stations. A great variety of signaling methods are theoretically possible. Some of these systems are described shortly. Since electromagnetic waves travel with a speed of about ~186,000 miles (299,274 kilometers)/second, the difference between arrival times of the signals will be very small. The unit of time used in these systems is the microsecond (0.000001 second); a difference between arrivals of one of these units indicates a distance difference of 0.186 miles (0.299 kilometers) from the two transmitters.

Points of constant difference of time between arrival of the two signals will fall on spherical hyperbolas, with the transmitters at the foci. For navigational purposes, one need only consider the lines of intersection of these surfaces with the surface of the earth. The total number of distinguishable lines in any system is equal to the time required for the signal to travel from the master to the slave and back again, divided by the smallest time interval that can be measured by the receiving equipment.

In some systems, two or more slave stations may be used with a single master. The cycle of transmission always begins at the master station, and the signal travels out in all directions. The arrival of the master signal at the slave "triggers off" the slave. Operators continuously monitor both stations, each monitoring the signals for the other, to detect the slightest variations of frequency carrier waves, intervals between pulses, and characteristics of the signals. At the slave station, adjustable delay circuits are available, so that, once the cycle has been started, the slave station may transmit simultaneously with the master or be delayed by any desired amount.

As shown in Fig. 8, certain fundamental lines, representing integral multiples of distance or time difference, are superimposed on regular navigational charts. Tables are published that contain the data for determining the fundamental lines. By graphical interpolation on the chart, or by mathematical interpolation from tables and stored data, the navigator can determine a hyperbolic line of position, using the observed difference in time of arrival of the signals, and the particular stations being used. The accuracy of the line varies from about 200 yards (183 meters) up to about 2 miles (3.2 kilometers), depending upon the distance of the observer from the base line between stations, and the type of equipment in use.

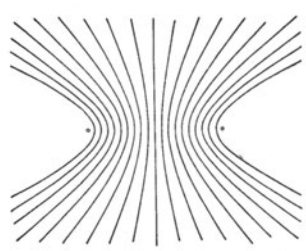


Fig. 8. Fundamental curves for hyperbolic navigation.

A fix as determined by a hyperbolic navigation system employing one master station, M, and two slaves, S_1 and S_2 (as in GEE navigation) is shown in Fig. 9. The diagram also indicates the value of hyperbolic navigation for "homing" on point A. The navigator obtains a fix at P and then sets his indicating equipment so that the pips from M and S_1 are in coincidence on the display instrument. Then the navigator heads the plane or ship so that, when these pips remain in coincidence, the ship will be following the hyperbola PA. By taking observations on the MS_2 pair at intervals, the navigator can determine the rate at which the objective is being approached.

DECCA Navigation. This is a system of hyperbolic navigation which employs low-frequency continuous-wave radiation.

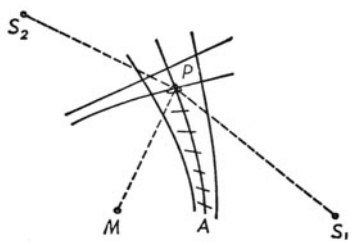


Fig. 9. Fix (hyperbolic navigation).

The master and any slave station radiate continuous waves, whose frequencies are related by a simple fraction, say one-fourth. The radiations from the two will be in phase when the distances from the stations differ by even multiples of a specified unit. This unit is a function of the wavelengths radiated.

In practice, a master and two slave stations are used. The receiving set reduces all three to a common frequency, and phase meters indicate the relative phase of each slave to the master. The accuracy of the setting of the phase meter is such that differences in distance may be determined with an accuracy of the order of magnitude of 100 feet (30.5 meters), and this is independent of distance from the base line. However, there is complete ambiguity of position, since there is no positive method of determining the number of complete phase changes between the observer and either station.

In spite of the ambiguity mentioned above, the system is of great accuracy and value when used in proceeding to some specified objective. A line of position, involving the master and one of the slave stations, is selected, which passes through the desired objective. The pilot must get the ship onto this line and then set his phase meter. If the pilot then proceeds so that the phase meter setting remains constant, the craft must be following the hyperbola directly to the objective. The hyperbolic lines from the master and the other slave will intersect the hyperbolic track along which the ship is proceeding. The pilot computes the number of complete phase changes that are to be expected, between the point of departure and the objective, along the line that the craft is following. When this number, plus any remaining fractional phase change, has been completed, the pilot must be directly over the objective.

LORAN Navigation. This is a long-distance radio-navigation system for aircraft and ships, utilizing synchronized pulses transmitted simultaneously by widely-spaced transmitting stations. Hyperbolic lines of position are determined by measuring the difference in the time of arrival of these pulses. The intersection of two of these lines of position, obtained from either three or four stations, gives a position fix. Standard LORAN operates on frequencies between 1,800 and 2,000 kHz. LORAN C is a widely used version of LORAN that uses pulse signals for more precise time-delay measurement and operates at a frequency of 100kHz. The range is 2,000 nautical miles (3706 kilometers). LORAN D is a tactical LORAN system that uses the coordinate converter of low-frequency LORAN C.

In standard LORAN, the time systems of the master and slave stations are such that the signal from the master always reaches a ship during the first half of the recurrence cycle, and that from the slave during the second half. This is accomplished by including a delay circuit in the slave timing system that delays the retransmission of the signal received from the master until half the recurrence period has elapsed.

Standard receiving equipment has been designed for ships and planes in which both the receiving and timing units are present, with selector switches permitting the operator to set on the frequency and recurrence rate assigned to any LORAN pair he wishes to use. Differential amplifiers, synchronized by the timing circuit to the recurrence rate of the station, act on both the master and slave signals to deliver them at equal strength to the indicator unit.

The slow sweep of the oscilloscope ("viewing scope") appears as two parallel lines, one covering the first half of the recurrence cycle and the other the second half. Hence the signals received from the master appear on one line and from the slave on the adjacent parallel line (see Fig. 10. (c)). An adjustment is provided to allow for correction of slight variations in the crystal control of the timing unit and, when this is properly set, the desired signals remain stationary on the scope. The signals from other stations, which may be within range and operating on the same frequency, will drift along the line since their recurrence interval will be different from that of the pair being used.

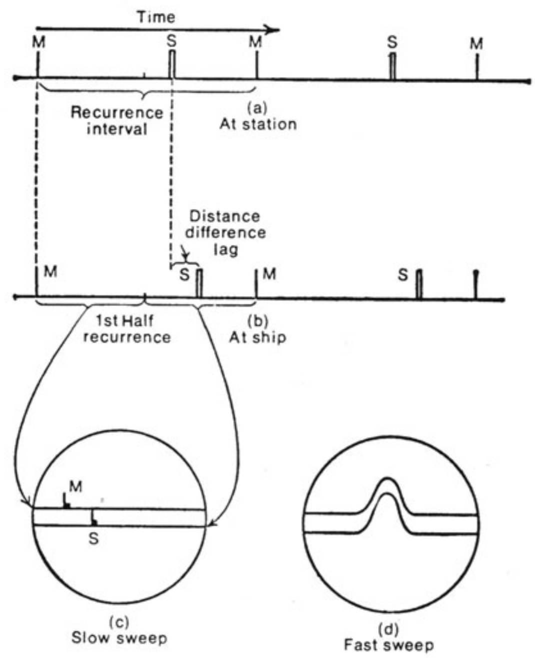


Fig. 10. Characteristics of LORAN signals.

When the signals are properly "set up" on the scope, a set of time markers is thrown on the screen, and a determination is made of the time interval between reception of signals from the master and slave. This will include the delay interval at the slave station, but, since this is standard for each recurrence rate, it may be allowed for. A delay circuit is now introduced, and the signals brought into approximate coincidence. With this adjustment made, a fast sweep spreads out the signals so that close coincidence may be established (see Fig. 10(d)), and the time interval measured to within one microsecond.

Using this measured difference in time of arrival of the two signals, the navigator then uses either tables or LORAN charts to obtain one line of position. The selector switches are then set to the characteristics of another LORAN pair, a second line of position determined, and a fix obtained (see Fig. 11). The accuracy of the determined fix is of the same order of magnitude as that obtained from good celestial navigation and, of course, is independent of the state of visibility.

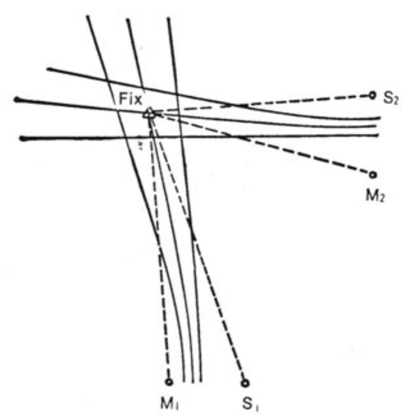


Fig. 11. Obtaining a fix with LORAN.

LODAR. This is a direction finder with which the direction of arrival of LORAN signals is determined free of night effect by observing the separately distinguishable ground and skywave LORAN signals on a cathode ray oscilloscope and positioning a loop antenna so as to obtain a null indication of the component selected to be most suitable.

OMEGA Navigation This is a long-range system that can provide worldwide coverage with only eight stations. The operating frequency is 10 kHz, and the estimated accuracy is 5 nautical miles (9.3 kilometers) with typical receivers. The system was developed for totally submerged submarines. The operating principle is similar to *delrac*, a British radio-navigation system designed to provide worldwide coverage by using 21 pairs of master-slave stations, with a 3,000-mile (4827 kilometers) range for each pair of stations. Frequencies used in *delrac* are in the band from 10 to 14 kHz. DECCA indicating equipment can be used with *delrac*.

GEE Navigation This is a vhf (very high frequency) radio navigation system developed in Great Britain and is similar to LORAN. For the transmission of the signals, one master and two or more slave stations are used. The distance between stations is about 75 miles (121 kilometers) and the stations are located approximately on a circle, with the master station between the slaves. The frequencies used are between 20 and 88 MHz, and the length of the pulses are of the order of magnitude of 6 microseconds. The accuracy of the lines of position varies with the square of the distance of the ship from the base lines between stations. On this line, the accuracy is of the order of magnitude of 200 yards (183 meters) when the navigator is on the base line, and about one mile at a distance of 400 miles (644 kilometers) from the base. Since time differences can be read simultaneously from the master and two slave stations, the fix can be determined by simultaneous observations without the necessity of using the running fix method. The system is excellent for "homing" on a particular objective.

TACAN System An air-navigation system in which a single uhf (ultra high frequency) transmitter sends out signals that actuate airborne equipment to provide range and bearing indications with respect to the transmitter location when interrogated by a transmitter in the aircraft. Each TACAN station broadcasts a location-identifying Morse-code signal at regular intervals. Also termed tactical air navigation.

SHORAN System A precision short-range position-fixing system using a pulse transmitter and receiver in an aircraft or other vehicle and two transponder beacons at fixed points. A receiver in the aircraft measures the round-trip times of the signals and converts these into distances to the fixed ground stations. Ordinary triangulation on a map then gives position.

OMNI System. This is a radio system that includes the directional information by modulating its radio signal with a simple dot-and-dash code, one code for a position to the left of the beam and another for a position to the right of the beam. The stations are located on airways and at the approaches to airways and to airports. With the system, the pilot selects his OMNI way point or destination on the radio and listens for the code that tells when the aircraft is to the left or right of the path. The system has been highly refined with onboard computers and displays. The system also is supplemented with distance-measuring equipment (DME). With the latter, the pilot interrogates the station with a transmitter and receives a distance indication from the station on a receiver. The measurement is made by determining the travel time of the radio wave. DME complicated the airborne equipment considerably but opened the way for easy, continuous navigation by using only a single set of equipment. Use of two stations can provide coverage for all locations within their range and thus free the pilot from flying the designated lines toward the OMNI stations. Very-high-frequency omnirange operates in the band from 112 to 118 MHz.

Doppler Navigation System This is a navigation system for aircraft which makes use of the doppler effect as a means for determining drift and ground speed. In one configuration, there are four beams of pulsed microwave energy which are beamed toward the ground (along the corners of an imaginary pyramid). The peak of the pyramid is at the aircraft. The echoes from the front-pointing beams experience an upward doppler shift, whereas the echoes from the rearward beams experience a downward doppler shift. Any drift is determined by doppler shift of echoes from beams on either side of the aircraft. The doppler shifts are compared in a computerized system, enabling all necessary navigation information under adverse weather conditions, various altitudes, and with need for reference to ground stations.

In a navigation satellite system, the satellite transmits accurate time signals and position data to a receiver on board a ship or aircraft. A central ground tracking station transmits correction signals to the satellite many times each day to sustain high accuracy of the system. The objective of such satellites, from which radio doppler shift measurements can be made under all weather conditions, is to provide the position of a ship or aircraft anywhere on earth with an accuracy of about 0.5 nautical mile (0.93 kilometer) or better.

NOTE: Numerous other articles in this encyclopedia relate directly or indirectly in navigation. Check alphabetical index.

Additional Reading

Baker, D. J.: "Toward a Global Ocean Observing System," *Oceanus*, 76 (Spring 1991).

Beardsley, T.: "Messages from on High," *Sci. Amer*., 112 (July 1988).

Bjerklie, D.: "The Electronic Transformation of Maps," *Technology Review (MIT)*, 54 (April 1989).

Carron, M. J., and K. A. Countryman: "Developing Oceanographic Products to Support Navy Operations," *Oceanus*, 67 (Winter 1990/91).

Frye, D. W., Owens, W. B., and J. R. Vales: "Ocean Data Telemetry," *Oceanus*, 46 (Spring 1991).

Garver, J. G., Jr.: "A Love Affair with Maps," *Nat'l. Geographic*, 130 (November 1990).

Grosvenor, G. M.: "New Atlas Explores a Changing World," *Nat'l. Geographic*, 126 (November 1990).

Kiernan, V.: "Guidance from Above in the Gulf War," *Science*, 1012 (March 1, 1992).

Koehr, J. E.: "The United States Navy's Role in Navigation and Charting," *Oceanus*, 82 (Winter 1990/91).

Marden, L.: "Tracking Columbus Across the Atlantic," *Nat'l. Geographic*, 572 (November 1986).

McVey, V.: "The Sierra Club Wayfinding Book," Little, Brown and Co., Boston, Massachusetts, 1989.

Monastersky, R.: "Satellite Secrecy Doesn't Sink Scientists," *Science News*, 358 (December 8, 1990).

Richardson, P. L., and R. A. Goldsmith: "The Columbus Landfall: Voyage Track Corrected for Winds and Currents," *Oceanus*, 2 (Fall 1987).

Turbank, G.: "A New Way to Find Your Way," *Amer. Forests*, 10 (August 1989).

Waldrop, M. M.: "Flying the Electric Skies," *Science*, 1532 (June 30, 1989).

Wood, D.: "The Power of Maps," *Sci. Amer*., 88 (May 1993).

NAVIGATORS' STARS. A list of 55 stars has been designated as the "navigators' stars." The list was selected to cover the entire celestial sphere in such a manner that a navigator, no matter at what season or in what part of the earth he may be operating, will have two or three navigators' stars available for observation. The names and positions of these stars are listed in the Air Almanac, published by the U.S. Naval Observatory, and in a number of other publications. See **Celestial Sphere and Astronomical Triangle;** and **Navigation.**

NAZCA PLATE. See **Ocean.**

NEARSIGHTEDNESS. See **Vision and the Eye.**

NEBULA. The term nebula (*stella nebulosa*) was originally used by astronomers to describe any luminous spot that remained fixed relative to the stars. Before the application of the telescope, probably the only objects to which the term applied where those that are now referred to as star clusters, although reference was made in the tenth century to the great spiral in Andromeda. Following the application of the telescope, many more nebulous objects were discovered. Originally, these were grouped into three classes: the diffuse nebulae (see accompanying figure); the planetary nebulae; and the spiral nebulae. Further research indicated that the spiral nebulae were very different from the other two. Application of the large telescopes proved that, in reality, they are groups of stars. Thus, for this class, the term nebula was dropped and the term galaxy or spiral galaxy used.

Some nebulae shine only by reflecting the light of stars contained in them ("reflection nebulae," which are dust in the environment of cooler, young stars), but others contain extremely hot stars whose radiation knocks electrons into high energy levels in the nebula gas, exciting gas atoms and allowing them to radiate as the electrons return to their

The Great Nebula in Orion (NGC 1976), with NGC 1977 below. (*Lick Observatory.*)

ground states, producing soft glows also called HI regions. Some nebulae are stellar birthplaces; inside them, fresh stars are being produced out of the nebular gas and dust. Other nebulae are debris of explosions marking disruptions of unstable stars, supernova remnants like the Crab Nebula.

With the tendency for a mixture of gas and dust to collect in clouds and condensations, obscuring clouds of interstellar material, known as dark nebulae, may collect. When one of these clouds is in the vicinity of a bright star, the intense radiation from the star will illuminate the cloud, and a bright diffuse nebula is observed. Studies of the spectra of these objects have shown that the light is made up of both reflected starlight and of radiation from the interstellar material. The character of the reflected light is similar to that from the nearby stars. The radiation from the nebulous material itself, however, is quite different in character from starlight. This nebular spectrum is the bright-line type, and is probably produced by the absorption of radiation from the star by the gas atoms, and then reradiation of this energy in frequencies characteristic of the gas atoms and their states of excitation and ionization. This hypothesis has been considered sound because the character of the nebular spectrum depends upon the spectral type of the nearby stars. If the star is hot (B-type), the nebular spectrum is rich with bright lines, but in the vicinity of a relatively cool star (A-type or later), the nebular spectrum is almost entirely that of reflected starlight.

When the spectra of the bright nebulae were first studied, two bright lines were observed that could not be identified with any known terrestrial or solar element. At one time, it was believed that the lines were due to some material that existed only in the nebulae and the name of an element Nebulium was coined. These lines turned out to be due to forbidden transitions of doubly ionized oxygen, [OIII], which, at that time, had not been produced in the laboratory because of the very low density required to prevent collisions from de-exciting the atoms.

In addition to the diffuse nebulae, there are the planetary nebulae (see photograph accompanying entry on **Lyra**), so-called because they have quite definite shapes and look more or less like planets when observed through a telescope. The general appearance of these objects, on detailed photographs, is that of shells of gaseous material. They are generally elliptical in form, and may appear to be made up of several elliptical shells having their axes at various angles to each other or as helices. Frequently, a very blue, and hence, a very hot star is observed at the center of the shell, which is a hot white dwarf, the core of the star whose envelope the nebula represents. A number of these stars prove to be very short (less than one hour) period binaries. The spectra of the planetaries is, in general, the same as that of the diffuse nebulae found in the vicinity of B-type stars. In some cases, where the central star can be studied, variations in its light have been found, and the nebular radiation is found to vary with that of the star. It is evident that the planetary nebulae are actually stars with very extended and attenuated atmospheres. Careful studies of the spectral lines from the planetaries indicate that the shell may be expanding. Expanding shells of gas have been observed around some novae, but the rate of expansion is far greater than that for the planetaries.

Since the advent of radio astronomy, a significant number of different kinds of molecules have been found in the interstellar medium, mostly in dark clouds of dust where light from stars cannot penetrate. However, the first interstellar molecule to be discovered was the radical CH, found accidentally in a star spectrum in 1937 with conventional earth-based optical instrumentation. The radicals CN and CH^+ were found with the Mount Wilson Observatory 100-inch (254-centimeter) optical telescope in 1939. The discovery of other interstellar molecules had to await the development of radio telescopes. The OH radical was the first to be added to the list by radio techniques in 1963. Since that time, and particularly since the early 1970s, a number of new interstellar molecules have been added to the list. These include NH_3, H_2O, H_2CO (formaldehyde), CO, HCN, HC_3N (cyanoacetylene) CH_3OH (methanol), CH_2O_2 (formic acid), CS, CH_3CN (methyl cyanide), SiO, HNC, CH_3C_2H (methyl acetylene), NH_2CHO (formamide), CH_3CHO (acetaldehyde), HNCO (isocyanic acid), H_2CS (thioformaldehyde), H_2S, H_2CNH (methylene amine), and SO (sulfur monoxide), the latter found in 1973.

At one time, some investigators felt that the most important process in interstellar molecule formation was the combination of neutral molecules by radiative association, that is, reaction with photon emission. However, because the chemical activation energy required for neutral molecule reactions acts as a barrier to formation of more complex molecules, it was agreed that the reaction rates would be too slow for interstellar molecule formation. In contrast, the majority of reactions between ions and neutral molecules do not have such a barrier. Heavier ions can be built from lighter ones. Exemplary reactions include: $H_2^+ + H_2 \rightarrow H_3^+ + H$; or $H_3^+ + CO \rightarrow HCO^+ + H_2$. Ions can recombine with electrons to form neutral molecules, thus completing the reaction pathway. Electron density in dense clouds is estimated at 10^{-7} cm^{-3}. Commencing with the foregoing reactions, the building of complex molecules in dense clouds may require the assumption that the source of ionization is the flux of cosmic rays at energies of 100 MeV and upwards. The reactions also require solid surfaces, like dust grains, on which to occur. Densities greater than 10^4 cm^{-3} also appear necessary. In addition to a chain of reactions commencing with hydrogen, other possible chains could commence from ionized helium, carbon, and oxygen. However, some investigators feel that the hydrogen reactions are probably basic and that the ion-molecule scheme probably would not take place unless HCO^+ is present in the dense cloud.

Theories of interstellar molecule formation include the theory of formation on dust grains and the ion-molecule theory. Estimates indicate that many of the simple diatomic molecules may be formed by collisions in space; the more complex ones, such as H_2CO_2, CH_3CN, and NH_3CO, may be formed on grains of interstellar dust in clouds where the concentration of hydrogen molecules (H_2) is $10^6/cm^3$ or higher. See also list of entries given in entry on **Astronomy.**

Steven N. Shore

NECROSIS. The local death of cells results in changes in the tissue known as necrosis. These consist of disintegration of the cellular structure with destruction of the nucleus and coagulation or liquefaction of the cytoplasm. The causes of necrosis include interference with the blood supply of a tissue physical injury, and deleterious actions by bacteria or their toxins.

NÉEL TEMPERATURE. The transition temperature for an antiferromagnetic material. Maximal values of magnetic susceptibility, specific heat, and thermal expansion coefficient occur at the Néel temperature. See also **Antiferromagnetism.**

NEGATRON. A term sometimes applied to the normally occurring negatively charged electron when it must be distinguished from a positron. In many parts of the world the name *negaton* is used instead of negatron. The word negatron is used in this encyclopedia wherever distinction is made between positively and negatively charged electrons.

NEGRO BUG (*Insecta, Hemiptera*). Small shining black bugs with a smooth convex upper surface. They resemble beetles superficially. Most of the abdomen is covered by a greatly enlarged sclerite of the thorax which also conceals most of the wings.

NEIGHBORHOOD OF A POINT. The interior of some bounded geometric figure (such as a square or circle in the plane) which contains the point. A neighborhood or a point on a line, plane, or surface is usually taken as the set of points within a stated distance of the point (e.g., an open interval on the line or the interior of a circle in the plane, with the point as center). One speaks of a property as holding *in the neighborhood of a point* if there exists a neighborhood of the point in which the property holds, or of a numerical quantity (e.g., curvature) depending on the nature of a curve or surface in the neighborhood of a point if the value of the quantity can be determined from knowledge of the portion of the curve or surface in an arbitrarily small neighborhood of the point. See also **Point.**

NEKTON. The portion of a population made up of animals which are capable of directive locomotion through a fluid medium. Usually applied only to aquatic animals, including the fishes, although flying creatures constitute a similar part of the terrestrial fauna and may be called an aerial nekton. See also **Ocean.**

NEMATIC LIQUID CRYSTALS. See **Liquid Crystals.**

NEMATODES. Of the phylum *Nemata* or *Nematoda*, these are roundworms or threadworms. They are abundant in fresh and salt water and in the soil; many are internal parasites of animals and plants. Some are parasitic in humans and the domestic animals and are important in relation to human welfare. The body of these worms lacks a spiny proboscis and is marked by slender longitudinal lines along the sides. These lateral lines follow the excretory tubes. There is a wide range of variations in the life cycle. To place *Nemata* in their proper perspective, it is in order to mention the other important phyla of worms: Flatworms (*Platyhelminthes*); ribbon worms (*Nemertea*); spiny-headed worms (*Acanthocephala*); hairworms (*Nematomorpha*); and segmented worms (*Annelida*). Reference to the entry on **Intestinal Nematodes** is suggested.

Nematode Damage to Food Crops

Nematodes are very important economic pests on food crops. Very few crops are immune to attacks of these creatures which inhabit the soil about the roots of plants. Nematode populations number into the millions and billions, in field crops, orchard operations, greenhouse (glasshouse) facilities, and truck and home gardens. Actually, nematodes have been one of the last of the major crop pests to be well understood, and aggressive research in the field only dates back some 50 to 60 years. Research progress was impeded mainly by difficulties in isolating and preparing the nematodes for detailed examination. Some authorities place nematology just about one-half century behind entomology, but much progress has been made during the past decade or two and thus the technological gap is narrowing. Among the challenges facing the nematologist today are: (1) Development of nematode-resistant varieties of crop plants; (2) cooperation with agricultural engineers in development of more effective means for applying nematicides; (3) education of farmers and large food producers on cultural methods for controlling nematodes, including fallowing and rotation of crops; (4) development of improved systemic nematocides as well as synthetic plant diffusates which stimulate early emergence of nematodes from dormancy; and (5) a continuing program of identifying yet undiscovered species. Continuing work on the classification and nomenclature of nematology also is important.

Although the importance of nematodes to food growth economics was relatively late in being appreciated, a knowledge of the existence of nematodes dates back to ancient times. It is believed that the Guinea worms mentioned in the Old Testament (*Numbers* 21:6–9) as "fiery serpents" were nematodes. Parasitic nematodes were alluded to by Aristotle as early as about 350 B.C. Free-living nematodes were observed in vinegar, and referred to as vinegar eels, as early as the mid-15th century. In the late 1700s, Linnaeus, Scopoli, Steinbuch, and Needham showed a causal relationship between the nematode *Anguina tritici* and the disease known as "cockles" of wheat and other cereal plants. In the late 1800s and early 1900s, Julius Kühn and associates in Germany intensively researched the sugar beet nematode and some authorities credit Kühn with the first use of soil fumigation. He used carbon disulfide on infested sugar beet fields. The life cycle of this pest (*Heterodera schachtii*) was ascertained, providing knowledge upon which effective cultural practices could be established.

Galls on the roots of cucumbers were noted by Berkeley in England as early as 1855. It is believed that the term root-knot nematode was first used in 1879 by Cornu. In 1887, root galls on coffee were described by Goeldi. Bastian, who wrote a monograph on the *Anguillulidae* in 1886, is considered the father of nematology. One of the first full texts on the subject, "Nematodes That Are Important for Agriculture," was authored by the Russian I. N. Filipjev in 1934. Outstanding early work was done by N. A. Cobb at the U. S. Department of Agriculture in the early 1900s. An excellent summary of the history of nematology from its beginnings to the 1960s is given in "Principles of Nematology," by Gerald Thorne, McGraw-Hill, New York, 1961. Additional and more recent references are listed at the end of this article.

Nature of Nematodes

These economic pests are found essentially wherever soil is found—from deserts and tropical areas to cold, high-altitude mountainous terrains. For example, in 1929 Thorne found several hundred specimens in soil samplings taken from Colorado mountain soils at levels of over 14,000 feet. There appears to be an almost infinite variety of nematodes, the greatest variations probably occurring in marine waters near shallow coastal beaches.

The typical nematode may be described as a slender, quite active animal that ranges from 0.2 to 10 millimeters in length, although the majority are less than 2 millimeters long. The body is usually cylindrical in shape, although several other forms are known, including pear- and lemon-shaped forms. The nematode body is covered with cuticle, a tough, flexible layer of material. In some species, the coating is marked or texturized, which helps the nematologist greatly in identification. However, many are not so conveniently marked, thus requiring detailed microscopic examination to yield identity. On the average, a nematode will undergo four moults in developing from egg to adult. During these stages, the body increases in diameter and length.

Control chemicals for use against nematodes are various fumigants, such as carbon bisulfide, chloropicrin, D-D, EDB, formaldehyde, Fumazone, hydrogen cyanide, methyl bromide, and Nemagon. These and other chemical materials either are banned or are subject to rigid control in some countries.

Nematodes are both *endo-* (inhabit and consume internal organs of host) and *ectiparasitic* (live on surface of host). They can be classified roughly in terms of the portions of the host plant which they prefer for habitation. Some of these include:

Bud and Leaf Nematodes (genus *Aphelenchoides*). These nematodes exhibit both endo- and ectoparasitism, a factor determined by Franklin in 1950. The endoparasites are found in leaves; the ectoparasites are found in plant crowns, leaf axils, or inflorescences, where these parts are protected by other folding tissues of the host plant. Nematodes of this type were discovered by Ormerod in England in 1889.

Bulb and Stem Nematodes (genus *Ditylenchus dispaci* (Kühn) Filipjev). There are over 30 species. The symptoms of their presence were first observed by Schwertz in 1855 on clover, oats, and rye, but the nematode causative agent was not revealed until further studies by Kühn in 1857 and others at later dates.

Burrowing Nematodes (genus *Radopholus* (Thorne)). These nematodes were named and classified in 1949. They are endoparasitic, attacking plant roots. They probably are found in tropical and subtropical regions, but have been found in cooler regions as well. *Radopholus similis* (Cobb, 1893) causes banana plant disease (*Musa sapientum*), first noted in Fiji (1890). Also noted on diseased coffee roots in Java (1898) and on diseased sugarcane roots in Hawaii (1907).

Cyst-Forming Nematodes (genus *Heterodera* (Schmidt, 1871)). These include the first of the important nematodes to be associated with plant diseases, namely, the sugar beet nematode (*Heterodera schachtii*) first observed by Schmidt. The female cuticle of this pest transforms into a light-to-dark brown, cystlike sac. This sac protects the eggs. These cysts are oval or spheroidal in shape and range from 0.4 to 0.8 millimeter in length. Inasmuch as the cyst material does not decompose readily, there are often great accumulations of these bodies in the soil. While older cysts will not have eggs, those of more recent years may contain as many as 600 eggs.

Root-Knot Nematodes (genus *Meloidogyne* (Goeldi)). At one time these were considered to be a single species. Five or more species were established by 1949. These pests are among the most economically important of the nematodes and include the coffee root-knot nematode *Meloidogyne exigua* (Goeldi 1892); the Japanese root-knot nematode, *M. japonica* (Treub, 1885); the northern root-knot nematode *M. hapla* (Chitwood, 1949a); the peanut (groundnut) root-knot nematode *M. arenaria* (Neal, 1889); the Thames root-knot nematode *M. arenaria thamesi* (Chitwood, 1952); the southern root-knot nematode *M. incognita* (Chitwood, 1949a); and the cotton root-knot nematode *M. incognita* var. *acrita* (Chitwood, 1949a), among others. The root-knot nematodes are often associated with various fungus diseases. Damage is caused by formation of galls on plant roots, causing stunting and wilting and, frequently, expiration of the plant if not controlled.

Root-Lesion Nematodes (genus *Pratylenchinae* (Thorne, 1949)). These have been described since the late 1920s. Because of openings in plant roots caused by the pests, bacteria and fungi may enter and thus these nematodes are often associated with serious diseases from these causes. The root-lesion nematodes cause openings or lesions rather than knots or galls as in the case of the genus *Meloidogyne*. There are about 10 major species and they infest a variety of very important crops, such as coffee, citrus, pineapple, potato, rice, and sugar beet, among many others.

Nematode Damage

Important diseases caused by nematodes affect many crops. Nematodes are found almost universally in the soil and at one time or other contribute to the damage, minor or major, to nearly all plants. There are however, numerous situations where nematode damage can be severe and even catastrophic to some crops if effective control measures are not taken. Only three examples are given here. Much more detail will be found in the Considine (1981) reference listed.

Citrus. A number of nematode species damage various citrus crops. Possibly the most serious situation occurs on citrus in Florida and is a condition known as the *spreading decline of citrus*. See Figs. 1–3. Caused by a burrowing nematode, the damage has ranged into the many millions of dollars. Experiments in treating diseased trees, however, point to a factor, still unknown, that also is active in causing spreading decline disease. In some experiments, destruction of the nematodes did not fully prevent spreading of the disease. Symptoms of the disease include stunted trees with subnormal foliage, small fruit, and retarded terminal growth. Wilting is excessive during dry, hot periods. There is also a reduction of young feeder roots. The term *spreading* derives from the fact that the nematodes spread out or migrate from one tree to the next. A study made by Suit and Ford in 1950 indicates that the advance is at the rate of 1.6 trees per year. The nematodes have been known to migrate under highways and railroad rights of way. Rather aggressive methods have had to be used in attempts to eradicate the disease in Florida, including systematic and frequent soil inspections and the destruction by burning of infested trees and planted areas; and the extensive use of very strong chemicals to an average depth of 12 inches (0.3 meter). The persistence of the nematodes is exemplified by the finding

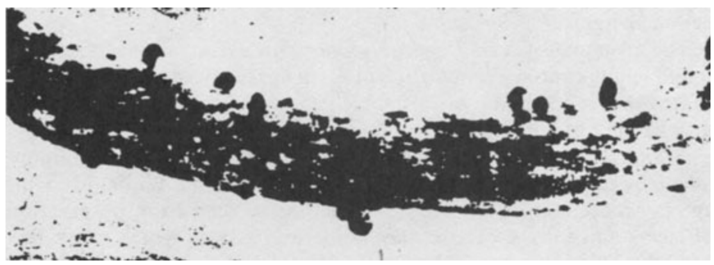


Fig. 1. Citrus root infested with mature female citrus nematodes. The actual diameter of the root shown is $\frac{1}{32}$ inch (0.8 mm). (*Agricultural Extension Service, University of California.*)

(a)

(b)

Fig. 2. Young Navel orange trees on Troyer citrange rootstock 2 years after planting: (a) trees planted in soil in which citrus nematodes had been killed by preplant soil fumigation; (b) tree infected with the citrus nematode soon after planting in nonfumigated, nematode-infested soil in same field as tree at left. (*Agricultural Extension Service, University of California.*)

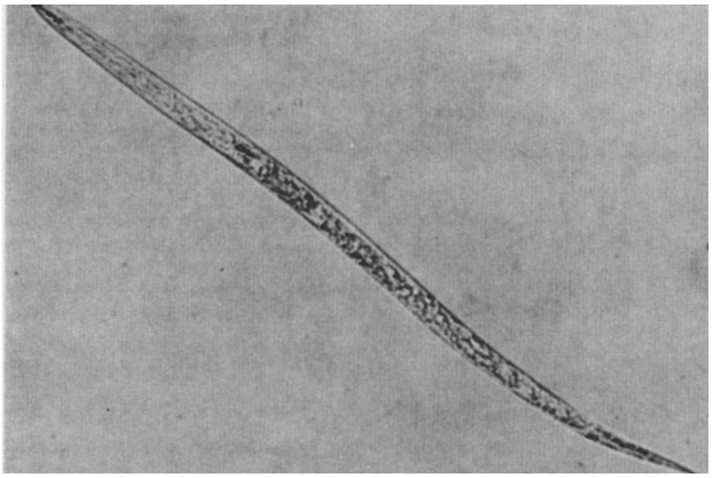


Fig. 3. Citrus nematode larva. Actual length is $\frac{1}{70}$ inch (0.4 mm). Stylet of feeding mechanism is at head (lower right-hand corner of view). (*Agricultural Extension Service, University of California.*)

of live nematodes within a depth of 12 feet (4 meters) of a very heavily treated area. However, inasmuch as most of the nematodes are found within the top 5 feet (1.5 meters) of the soil, diffusion of the treatment to a depth of 6 to 8 feet (2 to 2.5 meters) is usually effective.

Some authorities have observed a close parallel between the spreading decline of citrus disease in Florida and the yellows disease of pepper on Bangka.

The nematode species *Hemicycliophora arenaria* has been found on rough lemon root-stock in California. This pest causes an enlargement of terminal and lateral root-tips, which appear as small knobs. Damage may be the result of secretion of enzymes by the nematodes.

Potato. Nematodes of the genus *Ditylenchus* (Filipjev, 1934) injure potato by producing a progressive dry rot of the tuber. While this damage is proceeding, there is no evidence to be seen from observation of stems and foliage. There may be several strains which cause this condition. Known since 1888, *Ditylenchus destructor* has been a major cause of injury to potato in Europe for a long time. For many years, United States officials intercepted numerous shipments containing the strain.

Presence of the pest can be determined only by cutting into the tuber. In some instances much of the tuber can be infested. Crop rotation provides no effective control. Fumigation is the main control, and of the fumigants, ethylene dibromide is perhaps most effective. With sufficient fumigant applied, virtual eradication can be accomplished. Resistant varieties have not been successful. Where potatoes are used for silage, the fermentation of the silage kills the pests.

The golden nematode of potatoes, *Heterodera rostochiensis* (Wollenweber, 1923), is very damaging to potato. This nematode was discovered by Kühn in 1881 when he was doing research on the sugar beet nematode. Somewhat later reports were made in Germany and Scotland. The presence of the nematode in much of Europe was confirmed and, in the British Isles, the pest was called the potato root eelworm. During the interim, the nematode has been reported by Israel and, in 1941, the pest was first found in the United States on Long Island, New York. The nematode forms pear-shaped cysts on the potato, thus differing from the lemon-shaped cysts of the sugar beet nematode. This difference removed any doubt that the two pests are different. Since this was a new find in nematology and one affecting a huge market crop, research on the pest was intensive and many countries instituted crop regulatory and quarantine measures.

Rice. The rice stem nematode (*Ditylenchus angustus* (Filipjev, 1936)) is the cause of ufra disease in rice and was observed for the first time by Butler in 1913. This pest and resulting disease poses the greatest threat to success of the rice crop in India. It is found most commonly in an area north of the Bay of Bengal and east of the Ganges River. Even though many hundreds of varieties of rice are grown in these areas, all appear to be susceptible to this pest. Massive infestations in India have been known since 1916. The pest climbs the stems and interferes with the growing process of the plant, after which it consumes leaves and stems—in essence devastating the plant.

Additional Reading

Atkinson, H. J.: "The Physiology of Nematodes," 2nd edition, Columbia Univ. Press, New York, 1977.

Considine, D. M., Editor: "Foods and Food Production Encyclopedia," Van Nostrand Reinhold, New York, 1982.

Croll, N. A., Editor: "The Organization of Nematodes," Academic Press, Orlando, Florida, 1976.

Dropkin, V. H.: "Introduction to Plant Nematology," Wiley, New York, 1980.

Hollis, J. P.: "Action of Plant-Parasitic Nematodes on Their Hosts," *Nematologica*, **9**, 475–494 (1963).

Kontaxis, D. G.: "Nematicides Improve Sugar Beet Yields," *California Agriculture*, **31**, 4, 10–11 (1977).

Radewald, J. D., et al.: "Citrus Nematode Disease and Its Control," Agricultural Extension Service, Univ. of California, Berkeley, California. Bul. AXT-211 (revised periodically).

Rhode, R. A.: "The Nature of Resistance in Plants to Nematodes," *Phytopathology*, **55**, 1159–1162 (1965).

Stone, A. R., et al., Editors: "Concepts in Nematode Systematics," Academic Press, Orlando, Florida, 1983.

Thorne, Gerald: "Principles of Nematology," McGraw-Hill, New York, 1961.

Zuckerman, B. M., et al., Editors: "Plant Parasitic Nematodes: Morphology, Anatomy, Taxonomy, and Ecology," Academic Press, Orlando, Florida, 1981.

NEMERTEA. Marine worms with a flattened body, often long and ribbonlike in form. They are unsegmented and have no body cavity, hence they are sometimes included with the flatworms as a class of the phylum *Platyhelminthes*. More often they are made a separate phylum because the alimentary tract is a tube opening with an anterior mouth and a posterior anus. Like the free-living flatworms, they have ciliated (cilia) integument. They are also provided with an eversible proboscis enclosed in a dorsal tubular cavity, the rhynchocoel, and associated with but not derived from the alimentary tract.

These worms live among seaweed or at the bottom of the ocean and prey on living animals or eat dead ones. They are not economically important. A few freshwater species and a few parasitic forms are known.

NEODYMIUM. Chemical element symbol Nd, at. no. 60, at. wt. 144.24, third in the Lanthanide Series in the periodic table, mp 1,016°C, bp 3,068°C, density 7.004 g/cm^3 (20°C). Elemental neodymium has a close-packed hexagonal crystal structure at 25°C. The pure metallic neodymium is silver-gray in color, the luster becoming dull upon exposure to moist air at room temperatures. When pure, the metal is soft and malleable and may be worked with ordinary equipment. Because the metal is pyrophoric, it must be stored in an inert atmosphere or vacuum. There are seven natural isotopes, ^{142}Nd through ^{146}Nd, ^{148}Nd, and ^{150}Nd. ^{144}Nd is mildly radioactive with a half-life of $10^{10}-10^{15}$ years. Seven artificial isotopes have been produced. Of the light (or cerium-group) rare-earth metals, neodymium is the third most plentiful and ranks 60th in abundance of elements in the earth's crust, exceeding tantalum, mercury, bismuth, and the precious metals, excepting silver. The element was first identified by C. A. von Welsbach in 1885. Electronic configuration $1s^22s^22p^63s^23p^63d^{10}4s^24p^64d^{10}4f^35s^25p^65d^16s^2$. Ionic radius Nd^{3+} 0.995 Å. Metallic radius 1.821 Å. First ionization potential 5.49 eV; second 10.72 eV.

Other important physical properties of neodymium are given under **Rare-Earth Elements and Metals.**

Primary sources of the element are bastnasite and monazite which contain from 15 to 25% neodymium. Plant capacity involving liquid-liquid or solid-liquid organic ion-exchange processes for recovering the element is in excess of 200,000 pounds (90,720 kilograms) Nd_2O_3 annually. Metallic neodymium is obtained by electrolysis of fused anhydrous $NdCl_3$ or the electrolytic reduction of the oxide in molten NdF_3.

Use of elemental neodymium as a colorant for glass was one of the early applications. The color ranges from pure violet to purple and finds use in sunglasses, protective glasses for industry, art objects of glass, tableware, and decorative fiber optics. Use of neodymium in amounts of 3–5% by weight imparts dichroic properties to glass. Neodymium-doped single-crystal yttrium-aluminum oxide garnets (Nd:YAG) have been used in lasers. Research has shown the Nd ion to exhibit laser characteristics in a wide range of compounds and glasses.

A formulation of 75% neodymium and 25% praseodymium, frequently called didymium, is used as a metallurgical additive. Within the last several years, it has been found that the use of Nd_2O_3 in barium titanate capacitors increases the dielectric strength of these electronic components over a wider temperature range. Neodymium also has been used as an ingredient of phosphate-type phosphors. Investigations continue into further electronic and optical uses of the element and its compounds.

> EDITOR'S NOTE: Extensive research during the early 1980s led to the development of a new and powerful magnet material with the probable composition, $R_2Fe_{14}B$ (where R = a light rare earth). The rare earth predominantly used thus far is neodymium. The recent neodymium-iron-boron material exhibits extremely powerful magnetic qualities as compared with traditional magnet materials. More detail is given under **Rare-Earth Elements and Metals.** Also see **Magnetism.**.

Scientists (California Institute of Technology) reported that the isotopic composition of Drake Passage (Antarctica) seawater had been determined. The Antarctic Circumpolar Current, which controls interocean mixing, flows through the Drake Passage. The ratio, $^{143}Nd/^{144}Nd$, was found to be uniform with depth at two experiment stations—with an intermediate value between those of the Atlantic and Pacific Oceans. Further, Piepgras and Wasserburg determined that the Antarctic Circumpolar Current is made up of approximately 70% Atlantic water. It was further reported that cold bottom water from a site in the south-central Pacific has the Nd isotopic signature of the water in Drake Passage. The investigators used a box model to emulate the exchange of water between the Southern Ocean and ocean basis to the north with the isotopic results. An upper limit of about 33 million cubic meters/second was calculated for the rate of exchange between the Pacific and the Southern Ocean. Further determinations of samarium and neodymium were made and found to increase approximately linearly with depth. In essence, the findings suggest that Nd may be a valuable tracer in oceanography and possibly useful in paleo-oceanographic studies. See also **Ocean;** and **Polar Research.**

> 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.

Additional Reading

Anderson, D. L.: "Composition of the Earth," *Science*, 367 (January 20, 1989).

Cherfas, J.: "Proton Microbeam Probes the Elements," *Science*, 11500 (September 28, 1990).

Lugmair, G. W., et al.: "Samarium-146 in the Early Solar System: Evidence from Neodymium in the Allende Meteorite," *Science*, **222**, 1015–1017 (1983).

Letokhov, V. S.: Detecting Individual Atoms and Molecules with Lasers," *Sci. Amer.*, 54 (September 1988).

Lewis, R. J., Sr.: "Hawley's Condensed Chemical Dictionary," Van Nostrand Reinhold, New York, 1992.

Piepgras, D. J., and G. J. Wasserburg: "Isotopic Composition of Neodymium in Waters from the Drake Passage," *Science*, **217**, 207–214 (1982).

Robinson, A. L.: "Powerful New Magnet Material Found," *Science*, **223**, 920–922 (1984).

Staff: "ASM Handbook—Properties and Selection: Nonferrous Alloys and Pure Metals," ASM International, Materials Park, Ohio, 1990.

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

White, R. M.: "Opportunities in Magnetic Materials," *Science*, **229**, 11–15 (1985).

NEON. Chemical element symbol Ne, at. no. 10, at. wt. 20.183, periodic table group 18, mp −248.68°C, bp −246.01°C, density 1.204 g/cm^3 (liquid). Specific gravity compared with air is 0.674. Solid neon has a face-centered cubic crystal structure. At standard conditions, neon is a colorless, odorless gas and does not form stable compounds with any other element. Due to its low valence forces, neon does not form diatomic molecules, except in discharge tubes. It does form compounds under highly favorable conditions, as excitation in discharge tubes, or pressure in the presence of a powerful dipole. However, the compound-forming capabilities of neon, under any circumstances, appear to be far less than those of argon or krypton. No known hydrates have been identified, even at pressures up to 260 atmospheres. First ionization potential, 21.599 eV.

Neon occurs in the atmosphere to the extent of approximately 0.00182%. In terms of abundance, neon does not appear on lists of elements in the earth's crust because it does not exist in stable compounds. However, because of its limited solubility in H_2O, neon is found in seawater to the extent of approximately 1.5 tons per cubic mile (324 kilograms per cubic kilometer). Commercial neon is derived from air by liquefaction and fractional distillation. For most applications, the gas need not be in a highly pure form, but may be supplied along with small quantities of the other rare gases, such as argon and krypton. The gas finds principal applications in various electronic devices and lamps, but the most familiar application is the neon tubes used mainly in signs. The use of neon signs for identification and advertising signs reached the Iron Curtain countries at a date much later than in the Western countries. See *Smithsonian* magazine, **10**, 7, 150–154 (1979).) Neon emits the familiar orange light. Neon also has been used in certain lasers.

In the 1983 Luberoff reference, the author observes that neon signs, once considered vulgar symbols of a consumer society, are fast becoming icons of a bygone era. However, in recent years a group of preservationists, people who formerly decried the impact of neon advertising, now often defend it. Luberoff points out how a blue-lettered sign (5878 neon-filled glass tubes) became an integral part of the Boston skyline. A study of Boston's signs and lights by the Boston Redevelopment Authority showed that this sign (Citgo) was the only commercial sign that the public thought should remain.

There are three natural isotopes, ^{20}Ne through ^{22}Ne, and four radioactive isotopes, ^{18}Ne, ^{19}Ne, ^{23}Ne, and ^{24}Ne, all with half-lives of less than 5 minutes. Ramsay and Travers first found the element when investigating the properties of liquid air in 1898. The element is easily identified spectroscopically. Neon emits characteristic red and green lines in its spectrum.

Neon in Meteorites. As pointed out by Lewis and Anders, the noble gases are unique among the elements found in meteorites. They are highly volatile and unreactive and they did not condense in even the most primitive meteorites and thus are present at only a minute fraction of their proportion in the sun, ranging from about 10^{-5} for xenon to 10^{-9} for neon and helium. However, very small quantities of these gases are tightly bound in the meteorite and are freed when the host mineral begins to melt or decompose at high temperatures.

Scientists have found three types of neon in meteorites: (1) Primordial or planetary neon (called neon *A*); (2) solar neon (neon *B*), which consists of solar-wind neon ions implanted in meteorites that happen to have been at the surface of their parent body; and (3) cosmogenic neon (neon *S*), formed when cosmic rays passing through the meteorite spall, or shatter, atomic nuclei in their path. Each type has different proportions of the three isotopes of neon. Although the procedure is too detailed for inclusion here, Lewis and Anders explain how, through the use of stepped heating of meteorite materials, the types of neon can be measured and their ratios to each other provide clues as to what type of star may have been the source of a given meteorite.

Additional Reading

Anderson, D. L.: "Composition of the Earth," *Science*, 367 (January 20, 1989).

Cherfas, J.: "Proton Microbeam Probes the Elements," *Science*, 11500 (September 28, 1990).

Letokhov, V. S.: "Detecting Individual Atoms and Molecules with Lasers," *Sci. Amer.*, 54 (September 1988).

Lewis, R. J., Sr.: "Hawley's Condensed Chemical Dictionary," Van Nostrand Reinhold, New York, 1992.

Lewis, R. S., and E. Anders: "Interstellar Matter in Meteorites," *Sci. American*, **249**(2), 66–77 (1983).

Luberoff, D.: "But Is It Art? (Neon Signs), *Technology Review* (MIT), **86**(5), 76–77 (July 1983).

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

NEONATAL. Related to or affecting the newborn human child, particularly during the first month after birth.

NEOPLASM. Any new or abnormal overgrowth of cellular tissue. A neoplasm is a cellular tumor and may be either benign or malignant.

NEOPRENE. See **Elastomers.**

NEPHELINE. Nepheline, of hexagonal crystallization, is a sodium-potassium aluminum silicate (Na, K)($AlSiO_4$). It is found in silica-poor geological environments, where there had been insufficient silica to form feldspar. Nepheline rocks are characterized by the absence of quartz within them. They constitute a mineral family group known as the *feldspathoids*. Crystals are extremely rare; usually occurs massive to compact. Luster, is greasy in the massive varieties; vitreous in crystals. Color grades from yellowish to colorless in crystals; gray, green and reddish in massive material. It ranges from transparent to translucent. Hardness is of 5.5–6, specific gravity of 2.55–2.65.

Immense masses of nepheline-rich rocks occur on the Kola Peninsula, the former U.S.S.R., in Norway and in the Republic of South Africa; also in the Bancroft, Ontario, Canada region. Smaller deposits are found in Maine and Arkansas in the United States. Fine crystals are found in lavas on Mt. Vesuvius, Italy.

Nepheline is used extensively in the manufacture of glass.

Elmer B. Rowley, F.M.S.A., formerly Mineral Curator,
Department of Civil Engineering, Union College,
Schenectady, New York.

NEPHELOMETRY. Sir John Tyndall noted that particles which are invisible when directly in the path of a strong light become discernible when viewed from the side. Now known as the Tyndall effect, the phenomenon derives from reflection of part of the incident light by the particles. The reflected light is directly proportional to the number of particles in suspension. An instrument for measuring the intensity of reflected light so produced in a nephelometer and may be used for the quantitative determination of small amounts of diverse materials which have the ability to reflect light when in liquid suspension. Examples include the measurement of traces of silver wherein the chloride ion is added to a solution of material containing silver to produce insoluble silver chloride in suspension form. Small amounts of calcium in titanium alloys may be determined by measuring suspensions of the stearate formed in a suitable medium. Nephelometry also finds application in the measurement of bacterial growth rates, for the analysis of cholesterol, glycogen, and enzymes, for controlling the clarity of beverages, water, and wastewater, for solution control in tanning operations, and for any measurement situation where an unknown composition may be transformed into or related to a form of suspension.

Nephelometric methods are similar to fluorometric methods in that both involve measurement of scattered light. However, the scattering is inelastic in nephelometry and elastic in fluorometry. Thus, the scattered light which is measured in fluorometry is of a longer wavelength than the incident light, and both incident and scattered light are of the same wavelength in a nephelometric determination. In fact, the two functions sometimes are combined into one instrument, which may be termed a nefluoro-photometer. When the instrument operates as a nephelometer, it utilizes two Tyndall windows, located opposite each other in a cylindrical sample cell and with their common axis perpendicular to the path of the entering light. The concentration of suspended particles is determined by summing the photocurrents of the two cells. When used as a fluorometer, the instrument measures light emitted by a sample which is excited by incident radiation in the appropriate spectral band. Further, the same instrument can be set up for use as a photometer to measure light transmitted by the sample. Three light sources may be used—an incandescent source for colorimetric or nephelometric applications; a mercury-arc source for fluorometry; and a sodium-arc source with principal emission at 320 and 590 nanometers, when a sharp peak at either of these wavelengths is required, as in the instance of vitamin A determinations.

See also **Analysis (Chemical); Fluorometers; Photometers;** and **Turbidimetry.**

NEPHOMETER. See **Precipitation and Hydrometeors.**

NEPHRITIS. See **Kidney and Urinary Tract.**

NEPHRON. See **Kidney and Urinary Tract.**

NEPTUNE (Planet). Eighth planet from the sun (~3 billion miles; 4.5 billion km), Neptune is about four times the size (diameter) of Earth, with a mass slightly greater than 17 times that of Earth. Its mean density is less than that of Earth. This low density, coupled with the high value of the planet's albedo (0.52), is indicative of the planet's thick atmospheric layer. Neptune has seven confirmed satellites, the best understood of which is Triton.

Neptune is invisible to the naked eye, having a stellar magnitude of only about 7.7. The planet can be observed with a telescope having an aperture greater than 2.5 cm (1 in), but it can be distinguished with such a small instrument only by its change in position against the starry background from night to night.

The planet was discovered in 1846 and will not complete one trip around the sun from its position at that time until the year 2010, a total of 164 years later. Viewed by a large telescope, Neptune appears as a small, circular, somewhat greenish disk without distinctive markings. Thus, for many years, it was difficult to estimate the planet's rotation period. However, in 1928, Moore and Menzel (Lick Observatory) found, from spectroscopic observations employing the Doppler principle, that the planet rotates in the same directional sense as most other members of the solar system and with a rotation period of about 16 hours.

Voyager 2 Encounter with Neptune

The Voyager 2 spacecraft, initially conceived to visit Jupiter (July 9, 1981) and Saturn (August 26, 1981), was launched from Cape Canaveral, Florida on August 20, 1977, and performed so well that its mission was extended to include later encounters with Uranus (January 24, 1986) and Neptune (August 25, 1987). The dates given are for closest approaches.

Upgrading of System Instrumentation. With the pending investigation of Neptune in mind, after completion of the encounter with Uranus, project managers took advantage of the available time (approximately 3 years) to improve ground system instrumentation as much as possible to upgrade the data return for the Neptune encounter. These efforts included:[1]

- *Attitude control* was altered to reduce angular rates by approximately 25% below those experienced at Uranus, in an attempt to improve compensation for impulses from tape recorder starts and stops, to permit an additional scan platform rate useful for motion compensation near the closest approaches to Neptune and Triton, and to provide for "nodding" image motion compensation (NIMC). This permitted acquisition of motion-compensated images without disrupting the communication link with Earth or utilizing limited tape recorder resources. Instrument control for the imaging system also was changed to permit exposure durations between 15 and 96 seconds and real-time images with exposure durations in multiples of 48 seconds—capabilities that were used extensively during the Neptune encounter, where light levels were only 40% of those at Uranus.
- *Signal strength* is notably less from Neptune than from Uranus because of the great transmission distance between the spacecraft and Earth-based tracking systems. Each of the three 64-meter–diameter tracking antennas of the National Aeronautics and Space Administration (NASA) Deep Space Network (DSN) were enlarged to 70-meter diameter and improved in shape. A high-efficiency, 34-meter tracking station was added to the Madrid DSN complex. Extensive arraying of antennas was used to further increase the effective collector area. The 64-meter Parkes Radio Telescope in Australia again was made available to enhance the capability of the Canberra DSN tracking antennas. Similarly, Voyager signals collected by the National Radio Astronomy Observatory's Very Large Array (VLA) near Socorro, New Mexico, were combined with signals collected at the Goldstone, California DSN complex. The 27 25-meter antennas of the VLA provided a collecting area equivalent to 2 70-meter antennas. During closest-approach operations on August 25, 1989, the Japanese Institute of Astronautical Science utilized its 64-meter tracking antenna at Usuda, Japan, to augment Voyager science

[1]As reported by E. C. Stone (California Institute of Technology) and E. D. Miner (Jet Propulsion Laboratory).

TABLE 1. VOYAGER 2 INVESTIGATIONS AND MANAGERS

Investigation	Principal investigator and affiliation
Imaging (ISS)	B. A. Smith, University of Arizona, Tucson, AZ
Photopolarimetry (PPS)	A. L. Lane, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA
Infrared spectroscopy (IRIS)	B. J. Conrath, Goddard Space Flight Center, Greenbelt, MD
Ultraviolet spectroscopy (UVS)	A. L. Broadfoot, University of Arizona, Tucson, AZ
Radio science (RSS)	G. L. Tyler, Stanford University, Stanford, CA
Magnetometry (MAG)	N. F. Ness, Bartol Research Institute, University of Delaware, Newark, DE
Plasma (PLS)	J. W. Belcher, Massachusetts Institute of Technology, Cambridge, MA
Low-energy charged particles (LECP)	S. M. Krimigis, Applied Physics Laboratory, The Johns Hopkins University, Laurel, MD
Cosmic rays (CRS)	E. C. Stone, California Institute of Technology, Pasadena, CA
Plasma waves (PWS)	D. A. Gurnett, University of Iowa, Iowa City, IA
Planetary radio astronomy (PRA)	J. W. Warwick, Radiophysics, Inc., Boulder, CO

data collection. The spacecraft and all of the ground systems worked flawlessly during the Neptune encounter, testifying to the high level of expertise and teamwork within the Voyager project and its supporting organizations.

- *Trajectory design* was engineered to maximize the information returned during the Neptune-Triton encounters. The team had maximum freedom in this regard because no further missions for Voyager 2 were contemplated. Three primary objectives were sought: (1) a close approach to Triton, including both sun and Earth occultations as viewed from the spacecraft, (2) a close polar passage of Neptune, including both sun and Earth occultations, and (3) timing of the closest approach so that both of the Neptune-Triton occultations occurred at relatively high elevation angles over the Canberra DSN complex.

Voyager 2 Instrument Systems and Management. The investigatory disciplines and their management for the Neptune encounter are outlined in Table 1.

Chronology of the Neptune Encounter. Activities of the Voyager 2 encounter commenced on June 5, 1989, when the spacecraft was 117 $\times 10^6$ km from the center of the planet. The closest approach (center of Neptune) was a distance of 29,240 km and occurred on August 25, 1989. Radio signals from this position required a transit time to Earth of 4 hours, 6 minutes. The closest approach to Triton occurred on September 10, 1989, when the spacecraft was 39,800 km from the center of the satellite. This encounter period extended to October 2, 1989.

Design of the Neptune science sequences relied mainly on prior telescopic observations from Earth and on Voyager findings at Uranus, although some early Voyager data were used to make revisions to later observations. See Fig. 1. Provision was made for late retargeting of imaging frames to newly discovered satellites and rings. Timing of close Neptune and Triton observations also was adjusted at the last possible instant to take advantage of the most recent estimates of geometric event times made by the navigation team.

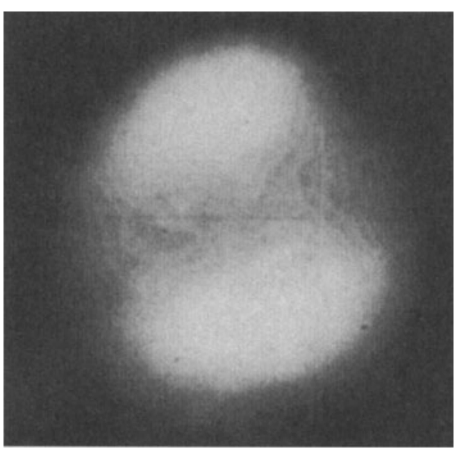


Fig. 1. Features of Neptune in which camera incorporating a charge-coupled device (CCD) was used with a 1.54-meter Catalina telescope. (*University of Arizona*, 1979.)

An initial, comprehensive report of the Voyager 2 encounter was released in December 1989. Detailed information is available from the Jet Propulsion Laboratory, Pasadena, California. Some of these reports were included in *Science* magazine, issue of December 15, 1989. Analysis of these data, including the formulation of explanations and postulations pertaining to the overall Neptunian system, will be forthcoming for several years into the future. A number of post-encounter papers issued during the early 1990s are listed at the end of this article.

The spacecraft's plutonium power sources may hold out until approximately 2015. Somewhat before that time, Voyager 2 is expected to encounter the heliopause (very edge of the solar system where the solar wind collides with interstellar media).

Atmosphere of Neptune

Images of Neptune were obtained by the narrow-angle camera of Voyager 2 and indicated large-scale cloud features that persist for several months or longer.[2] The periods of rotation of these features about the planetary axis range from 15.8 to 18.4 hours. The atmosphere equatorward of $-53°$ rotates with periods longer than the 16.05-hour period deduced from Voyager's planetary radio astronomy experiment. This is presumably the planet's internal rotation period. The wind speeds computed with respect to this radio period range from 20 meters per second eastward to 325 meters per second westward. Thus, it was found that the cloud-top wind speeds are approximately the same for all the planets ranging from Venus to Neptune, even though the solar energy inputs to the atmospheres vary by a factor of 1000.

Neptune has an effective temperature of about 59.3 K. Derivation of Neptune's Bond albedo continues to require a more thorough study of Voyager 2 instrument data. Neptune, however, appears to emit about 2.7 times as much energy as it absorbs from the sun. This greater contribution of internal heat may be the cause of the greater activity in the Neptunian atmosphere relative to that of Uranus. The horizontal temperature structures of the two atmospheres are quite similar, with the poles and equator at very nearly equal temperatures, while mid-latitudes are several degrees cooler. Temperature in the extreme upper atmosphere is nearly 750 K, but, because of Neptune's larger mass, colder atmosphere, and greater ring distances, the effects of gas drag on ring material are less at Neptune than at Uranus.

A number of prominent cloud features are apparent in images of Neptune's atmosphere, including an Earth-size "Great Dark Spot" (GDS) which occurs near $-10°$ latitude. The GDS is located and is of

[2]As reported by H. B. Hammel (Jet Propulsion Laboratory) and associated team members (see papers listed).

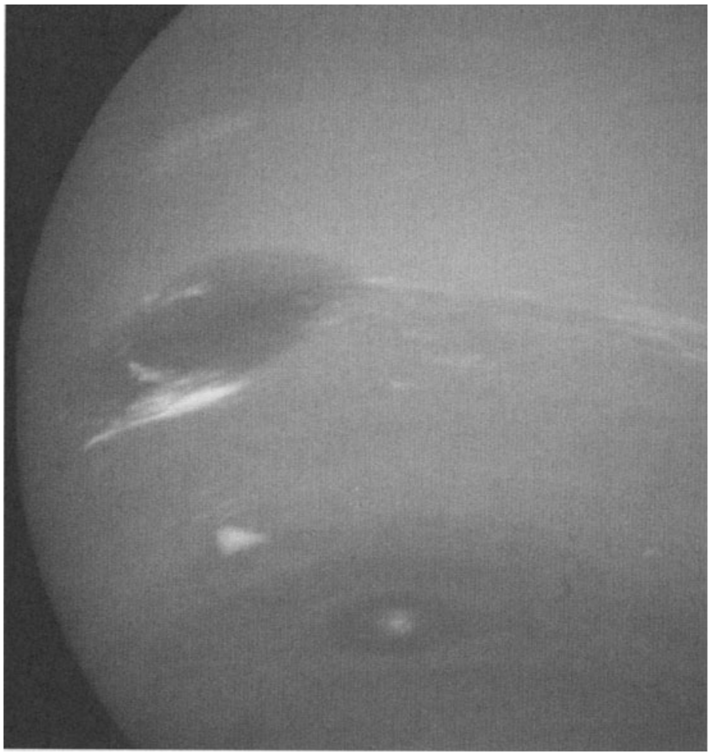


Fig. 2. Neptune's "Great Dark Spot," accompanied by white high-altitude clouds. (*National Aeronautics and Space Administration, Jet Propulsion Laboratory, Pasadena, California.*)

approximately the same size as Jupiter's Great Red Spot. See Fig. 2. The GDS rolls in a counterclockwise direction, with a 16-day period. Another, but smaller, dark spot with a bright central core is located at $-55°$ latitude. Bands of lower reflectivity extend from $+6°$ to $+25°$ latitude and from $-45°$ to $-70°$ latitude. The GDS is flanked by cirruslike cloud features. Other similar features occupy relatively narrow latitude ranges near latitudes of $-27°$ and $-71°$. These are believed to be optically thick upward extensions of the methane (CH_4) cloud deck. The details of these features change with time scales that are smaller as compared with the 16-hour rotation period. One of these features, referred to as the "Scooter," is bright and found near the $-42°$ latitude. It is deeper within the atmosphere than the aforementioned cirrus clouds and is postulated to be an upward extension of the deeper cloud deck. Velocities measured with respect to the internal rotation of Neptune indicate wind speeds ranging from about $+20$ m s^{-1} (prograde) at $-54°$ latitude to -325 m s^{-1} (retrograde) at $-22°$. The GDS resides in a region with strong wind shears and is believed to lie at a lower level than most of the brighter cloud features, virtually independent of the higher-altitude winds. The cirruslike clouds are found at altitudes of 50 to 100 km above the lower cloud layer. Optically thin layers of haze, believed to be produced photochemically from CH_4, are found at still higher altitudes.

As with the other giant planets, hydrogen (H_2) predominates the Neptunian atmosphere. Although subject to further calculations from measurements taken, the mole fraction of atmospheric helium [He]/[H_2] is estimated to be less than 0.25. Methane is more abundant in Neptune's upper atmosphere than in the atmosphere of Uranus. The absorption of red light by CH_4 gives Neptune its characteristic blue color. Deeper in the atmosphere, acetylene was detected. The signature of an optically thin cloud deck of methane ice was observed in radio occultation data. Strong absorption of radio waves may indicate the presence of small amounts of ammonia (NH_3).

Some further analyses of wind speed data are in order. Some scientists believe that the winds on Neptune are faster than found on any other planet, a surprising situation because of the small amount of energy received from the sun or from the interior of the planet. Past conceptual models of the general circulation of the giant planets do not readily provide a simple explanation as to why the highest winds appear to occur on Neptune, with its assumed low-energy sources. Neptune receives an estimated $\frac{1}{900}$ of the Earth's input of solar energy, but may have wind speeds of nearly 600 meters per second. How the near-supersonic winds can be maintained has puzzled some investigators. Scientists at the Space Science and Engineering Center of the University of Wisconsin–Madison have offered the following hypothesis: "Based on principles of angular momentum and energy conservation in conjunction with deep convection, leads to a regime of uniform angular momentum at low latitudes. In this model, the rapid retrograde winds observed are a manifestation of deep convection, and the high efficiency of the planet's heat engine is intrinsic from the room allowed at low latitudes for reversible processes, the high temperature at which heat is added to the atmosphere, and the low temperatures at which heat is extracted." (See Suomi/Limaye/ Johnson reference listed.)

A scientist (California Institute of Technology) observes, "Neptune's supersonic winds are not a certainty. The altitudes of the different cloud features, for example, are difficult to confirm, making it hard to compute the clouds' speeds. Moreover, it is difficult to tell from the photos whether the movements represent actual fluid motion of atmospheric masses or merely a wave moving through the atmosphere." (See Eberhart reference listed.)

Rings of Neptune

Earth-based stellar occultation measurements made during the early and middle 1980s alerted investigators to the probable existence of rings or, at least, partial rings (ring arcs) at a number of radial distances from the center of the planet. Voyager 2 imaging confirmed the presence of a system of at least six rings of prograde, equatorial, and circular rings. The outermost ring occurs at a distance of 62,900 km from the center of the planet. It is described as being composed of three bright, dusty areas. Data on Neptune's rings are given in Table 2.

Narrow rings are believed to be confined by the actions of relatively nearby satellites (shepherds). These rings may serve to prevent material from spiraling inward toward the planet. No ring shepherds have been noted thus far in the Neptunian system. The Voyager 2 instrumentation, however, was limited to observing satellites of a diameter of 12 km or

TABLE 2. PROPERTIES OF NEPTUNE'S RINGS

Feature	Distance (10^3 km)	Distance (R_N)	Width (km)	Optical Depth	Comments
1-bar atmosphere	24.76	1.000			Equatorial radius of Neptune
	38.	1.5		<0.0001	Inner extent of 1989N3R?
1989N3R	41.9	1.69	1700*	0.0001	High dust content
	49.	2.0		<0.0001	Outer extent of 1989N3R?
1989N2R	53.2	2.15	†	0.01	High dust content
1989N4R (inner)	53.2	2.15		0.0001	Inner edge of "plateau"
1989N4R (outer)	59.	2.4		0.0001	Outer edge of "plateau"
1989N1R	62.9	2.54	15	0.01–0.1	Contains three bright dusty arcs

*Tabulated width of 1989N3R is full width at half maximum. †1989N2R is narrow and unresolved in Voyager images.

Source: California Institute of Technology and Jet Propulsion Laboratory.

TABLE 3. PROPERTIES OF NEPTUNE'S SATELLITES

Satellite Name	Distance (10^3km)	Distance (R_N)	Period (Hours)	Diameter (km)	Resolution (km per Line Pair)	Normal Albedo
1989N6	48.0	1.94	7.1	54 ± 16	47.2	0.06?
1989N5	50.0	2.02	7.5	80 ± 16	34.8	0.06?
1989N3	52.5	2.12	8.0	180 ± 20	36.8	0.06?
1989N4	62.0	2.50	10.3	150 ± 30	33.8	0.054
1989N2	73.6	2.97	13.3	190 ± 20	8.2	0.056
1989N1	117.6	4.75	26.9	400 ± 20	2.6	0.060
Triton	354.8	14.33	141.0	2705 ± 6	0.8	0.6–0.9
Nereid	5513.4	222.65	8643.1	340 ± 50	86.6	0.14

Source: California Institute of Technology and Jet Propulsion Laboratory.

greater. Thus, tiny satellites may have escaped attention. Thus, it is not known whether or not additional shepherding satellites exist. Such material, if azimuthally unrestrained within the ring, should spread relatively uniformly around the ring within a time span of a few years. Additional encounters with largely enhanced resolution may be required at some future date to fully explain the dynamics of the planet's ring system.

Particles within the rings appear to be smaller than those found in the rings of Uranus. The dust content of one ring (1989N3R) is nearly double that of the other rings and thus compares better with that of the rings of Saturn and Uranus.

From analysis of Voyager 2 data, C. Porco (Department of Planetary Sciences, University of Arizona) proposes an interesting explanation for Neptune's ring arcs, "A radial distortion with an amplitude of approximately 30 km is traveling through the ring arcs, a perturbation attributable to the nearby satellite Galatea. Moreover, the arcs appear to be azimuthally confined by a resonant interaction with the same satellite, yielding a maximum spread in ring particle semimajor aces of 0.5 km and spread in forced eccentricities large enough to explain the arcs' 15-km radial widths." Additional ring arcs were discovered during the course of the study and provide further support to this model. (See Poroco reference listed.)

Triton and Other Satellites[3]

During its approach to Neptune, Voyager 2 images revealed six new satellites. All satellites orbit the planet in prograde, circular orbits of low inclination. Characteristics of the satellites are summarized in Table 3. Five of the six satellites orbit within 1° of Neptune's equatorial plane. The 1989N6 satellite has an inclination of nearly 6°. Data from Earth-based observations and Voyager 2 (see Figs. 3-10) show Triton's inclination to be 157°. Nereid's inclination is 29°. The respective orbital eccentricities of Triton and Nereid are 0.00 and 0.75. Nereid's distance from the center of the planet ranges from 1.39×10^6 to 9.64×10^6 km. Nereid's highly elliptical orbit makes it theoretically unlikely that its rotation and orbital periods are equal. However, Voyager 2 detected no rotational brightness variations in excess of 10%.

Triton. Much scientific interest has been directed toward this, by far the largest of Neptune's satellites and the existence of which has been known by earthbound observations for several years. Of all known natural bodies in the solar system, Triton has the lowest surface temperature (38 ±4 K). Triton's atmosphere is predominantly nitrogen (N_2), with the presence of CH_4 in the lower atmosphere. The surface pressure, as measured by the radio science instrumentation aboard Voyager 2, is 16 ±3 microbars. It is believed that a thermal inversion may exist in the lower 5 km of the atmosphere, and thus a tropopause altitude of 25 to 50 km is inferred. It is uncertain whether the clouds and haze layers observed in this region of the atmosphere result from simple condensation or from surface eruptions. The temperature at altitudes above 400 km is 95 ± 5 K. Atmospheric nitrogen is transported from the illuminated polar regions to the unilluminated polar regions.

The surface of Triton is that of a geologically young body and is devoid of heavily cratered terrain. The polar regions (south of latitude −15°) are covered with seasonal ice, believed to be N_2. Spring in Triton's southern region extends for several years—for example, in the present time frame, from about 1960 to the year 2000. Seasonal ice presents a slightly reddish tint believed to result from organic compounds photochemically produced by interaction between methane and nitrogen. Energetic particle bombardment also may assist in producing these reactions. The equatorial regions of Triton at most latitudes contain a thin layer of nitrogen frost. The layer appears as a bright, slightly blue coloration, a layer that does not obscure underlying topography. Observations indicate that northward of the equator there is a variety of terrains. Some scientists have referred to this topology as reminiscent of the "skin of a cantaloupe."

This terrain dominates the western (trailing) hemisphere of Triton and is believed to consist of a dense concentration of pits (dimples) that

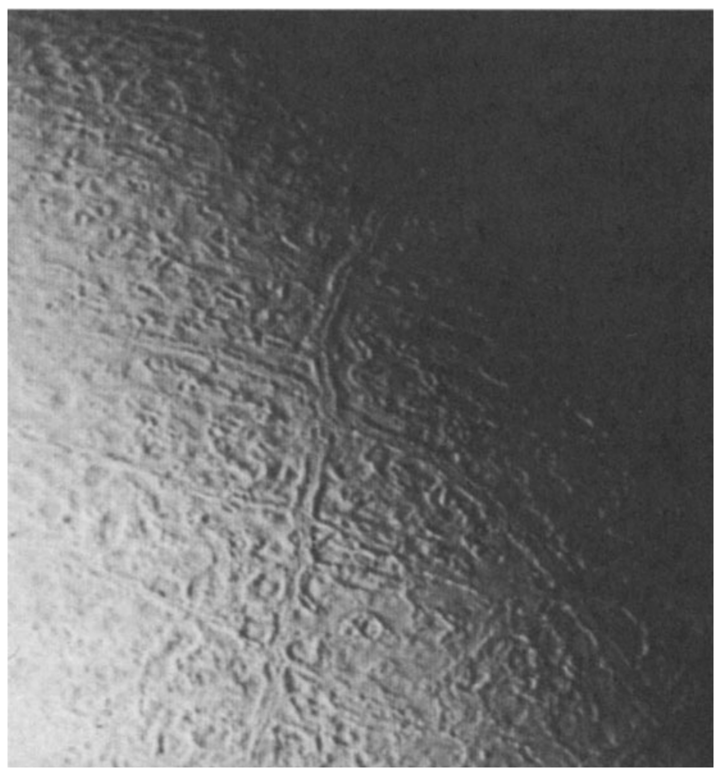


Fig. 3. Triton from 80,000 mi (128,720 km). Long feature is probably a narrow down-dropped fault. (*National Aeronautics and Space Administration, Jet Propulsion Laboratory, Pasadena, California.*)

[3]Principal information source: Initial report of Jet Propulsion Laboratory, Pasadena, California.

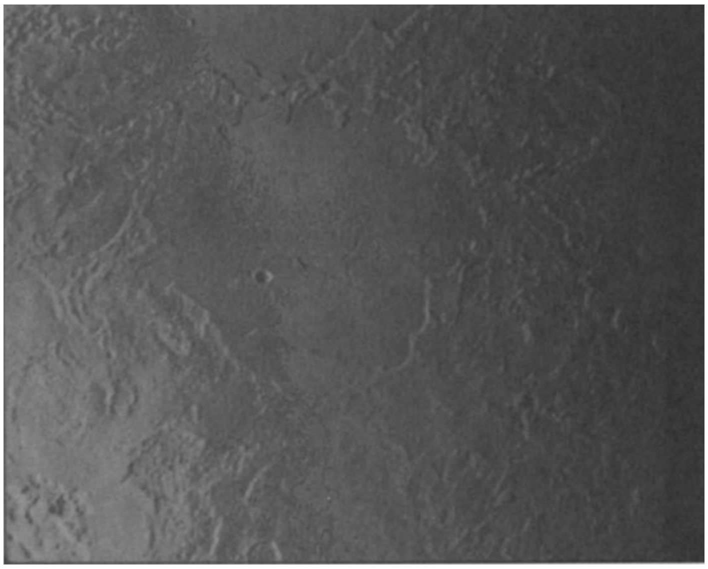


Fig. 4. View of about 300 mi (483 km) across Triton's surface. (*National Aeronautics and Space Administration, Jet Propulsion Laboratory, Pasadena, California.*)

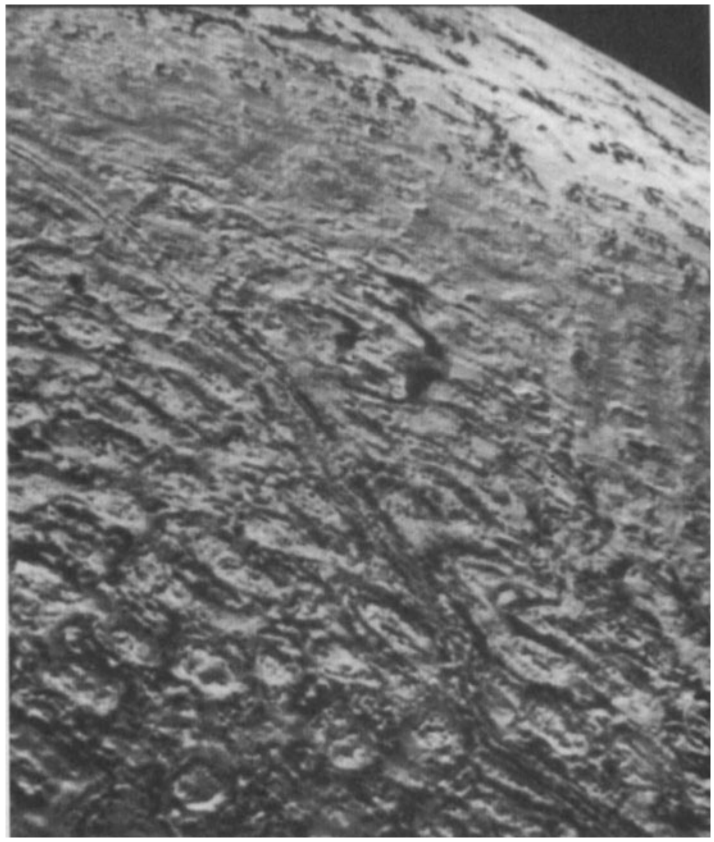


Fig. 6. Triton from 25,000 mi (40,225 km). Depressions may be caused by melting and collapsing of icy surface. (*National Aeronautics and Space Administration, Jet Propulsion Laboratory, Pasadena, California.*)

are crisscrossed by ridges. Few impact craters are discernible in these areas. The terrain of the eastern (leading) hemisphere of Triton is made up of a series of much smoother units, including caldera-like structures of water ice. Frozen lakes are surrounded by successive terraces, possibly due to a series of flooding actions.

Within the polar regions of Triton are numerous wind streaks with albedos that are 10–20% lower than the polar ices. The streaks, which overlie deeper ice deposits, appear to be young, possibly less than 1000 years of age. Two active geyserlike plumes were discovered near the subsolar latitude (−55°). As determined by stereoscopic viewing, these plumes rise to an altitude of about 8 km. Above a plume, dense clouds form and serve as a source for a westward wind-driven trail of material more than 100 km long. It has been suggested that the plumes may result from the explosive release of N_2 gas, which carries ice-entrained dark material in the exit nozzle to high altitudes.

In a report by L. A. Soderblom (U.S. Geological Survey, Flagstaff, Arizona) and a team of investigators, they explicitly describe the plume phenomenon: "The radii of the rising columns appear to be in the range of several tens of meters to a kilometer. One model for the mechanism to drive the plumes involves heating of nitrogen ice in a subsurface greenhouse environment; nitrogen gas pressurized by the solar heating explosively vents to the surface carrying clouds of ice and dark parti-

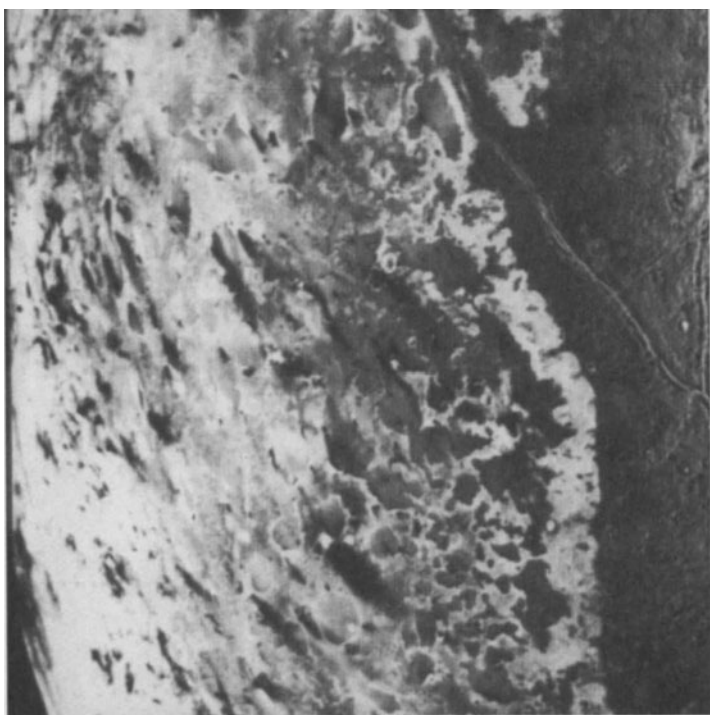


Fig. 5. Triton's south polar terrain. About 50 dark plumes mark what may be ice volcanoes. (*National Aeronautics and Space Administration, Jet Propulsion Laboratory, Pasadena, California.*)

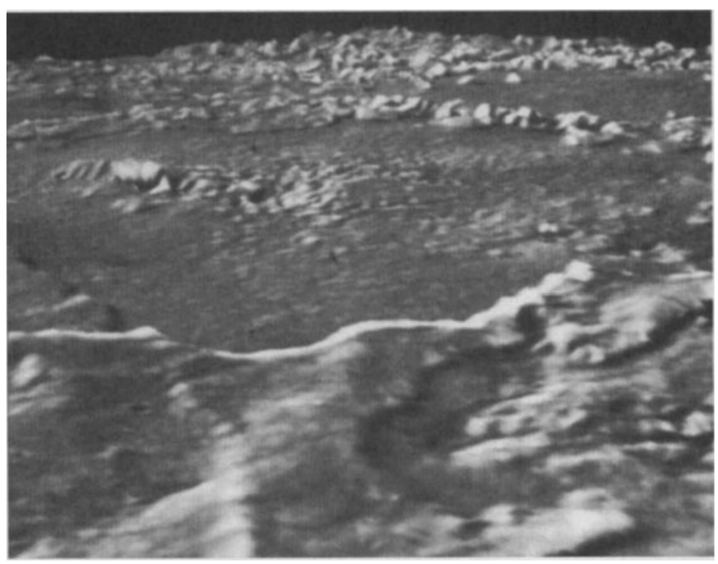


Fig. 7. Computer-generated perspective view of one of Triton's calderalike depressions. (*National Aeronautics and Space Administration, Jet Propulsion Laboratory, Pasadena, California.*)

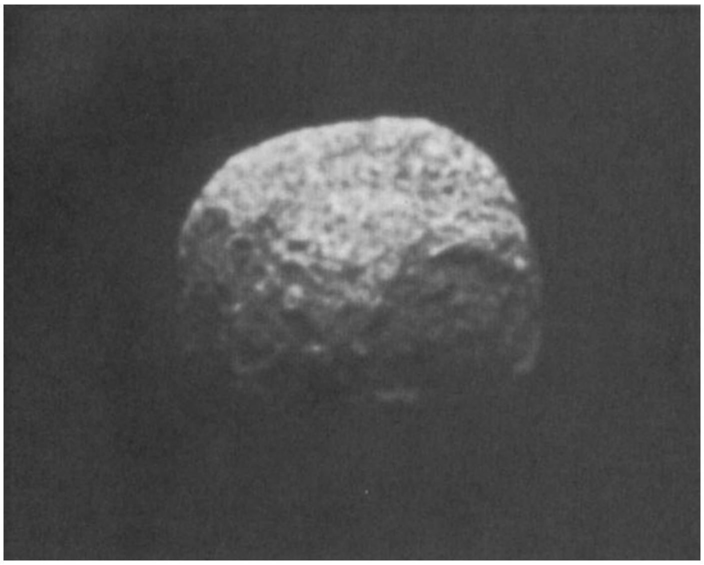


Fig. 8. Satellite 1989N1, discovered by Voyager 2. (*National Aeronautics and Space Administration, Jet Propulsion Laboratory, Pasadena, California.*)

cles into the atmosphere. A temperature increase of less than 4 kelvins above the ambient surface value of 38 ± 3 K is more than adequate to drive the plumes to an 8-km altitude. The mass flux in the trailing clouds is estimated to consist of up to 10 kg of fine dark particles per second, or twice as much nitrogen ice and perhaps several hundred or more kilograms of nitrogen gas per second. Each eruption may last a year or more, during which on the order of a tenth of a cubic kilometer of ice is sublimed."

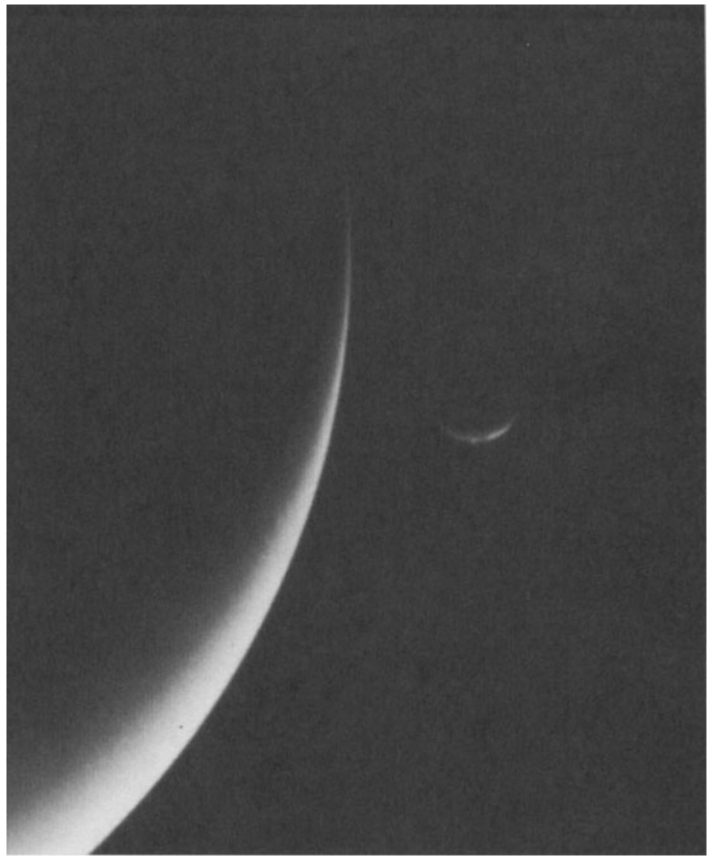


Fig. 10. Neptune and Triton 3 days after flyby. Triton is smaller crescent and is closer to viewer. (*National Aeronautics and Space Administration, Jet Propulsion Laboratory, Pasadena, California.*)

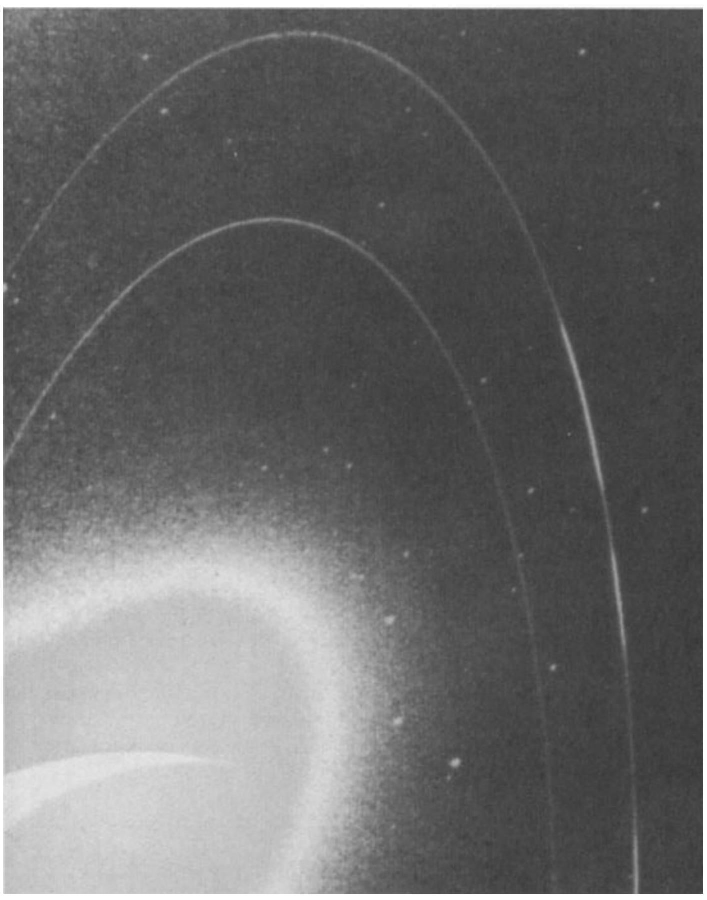


Fig. 9. Details of Neptune's rings. (*National Aeronautics and Space Administration, Jet Propulsion Laboratory, Pasadena, California.*)

In another approach to the plume phenomenon, A. P. Ingersol and K. A. Tryka (Division of Geological and Planetary Sciences, California Institute of Technology) observe, "Their structure suggests that the plumes are an atmospheric rather than a surface phenomenon. The closest terrestrial analogs may be dust devils, which are atmospheric vortices originating in the unstable layer close to the ground. Since Triton has such a low surface pressure, extremely unstable layers could develop during the day. Patches of unfrosted ground near the subsolar point could act as sites for dust devil formation because they heat up relative to the temperature of 48 K or higher, as observed by the Voyager radio science team. Assuming that velocity scales as the square root of temperature difference times the height of the mixed layers, a velocity of 20 m per second is derived from the strongest dust devils on Triton. Winds of this speed could raise particles provided they are a factor of 10^3 to 10^4 less cohesive than those on Earth."

Impact craters are rare on Triton—that is, those that are observable at the 3 to 1.8 km resolution acquired during the mapping sequence of Voyager 2. The highest-resolution images obtained show various degrees of smear, making analysis difficult. Thus, it is difficult to compare Neptunian cratering with other observed planets. The largest and uncontested impact crater viewed by Voyager 2 on Triton is only about 27 km in diameter. Several large quasi-circular features exist, but are believed to be of internal origin. Fresh impact craters on Triton have morphologies similar to those on other icy satellites seen at comparable resolutions. These features include simple sharp-rimmed and bowl-shaped interiors and a few craters with flat floors and central peaks. Based mainly on the similarity of size distribution on Triton and Miranda (satellite of Uranus) and the relatively young surface of Triton, comets are believed to be the primary source of cratering. On the other hand, the peculiar size distribution of sharp craters on the "cantaloupe" terrain and other evidence suggests that they are of volcanic origin.

Neptune's Magnetosphere

Eight days prior to Voyager 2's closest approach to Neptune, a distance of about 470 Neptune radii (R_N), the first indication of a Neptune magnetic field was obtained from radio emissions. Subsequently, the spacecraft crossed a well-defined, detached bow shock at 34.9 R_N. The inbound magnetopause was not as well defined because Voyager entered a highly tilted magnetic field at very high magnetic latitude, permitting the first observation of a "pole-on" magnetosphere in which the solar wind is incident on the magnetic polar region rather than the equatorial.

It has been determined that, as the magnetic field rotates with the planet each 16.11 hours, satellites and ring particles sweep through large ranges of magnetic latitude. The incident solar wind deforms the magnetic field, resulting in a well-developed magnetic tail behind the planet. However, as the planet rotates, the magnetosphere configuration changes from pole-on with a cylindrically shaped magnetotail plasma sheet to a more normal planar plasma sheet. Because of this unique geometry and the timing of the flyby, the spacecraft did not cross the plasma sheet.

No evidence exists for Neptunian electrostatic discharges of the kind observed on Saturn and Uranus by planetary radio astronomy (PRS). Many typical plasma waves were detected by the plasma wave instrumentation during the encounter. These included electron plasma oscillations in the solar wind upstream of the bow shock, and chorus, hiss, electron cyclotron waves, and upper hybrid resonance waves in the inner magnetosphere. There was no indication of lightning-generated whistlers.

See also **Voyager Missions to Jupiter and Saturn**.

Additional Reading

Belcher, J. W., et al.: "Plasma Observations Near Neptune: Initial Results from Voyager 2," *Science*, 1478 (December 15, 1989).
Broadfoot, A. I.: "Ultraviolet Spectrometer Observations of Neptune and Triton," *Science*, 1459 (December 15, 1989).
Brown, R. H., et al.: "Energy Sources for Triton's Geyser-Like Plumes," *Science*, 431 (October 19, 1990).
Brown, R. H., et al.: "Triton's Global Heat Budget," *Science*, 1465 (March 22, 1991).
Conrath, B., et al.: "Infrared Observations of the Neptunian System," *Science*, 1454 (December 15, 1989).
Cruickshank, D. P., et al.: "Triton: Do We See to the Surface?" *Science*, 283 (July 21, 1989).
Eberhart, J.: "Neptune Marvels Emerge from Data Deluge," *Science News*, 391 (December 16, 1989).
Goldreich, P., et al.: "Neptune's Story," *Science*, 500 (August 4, 1989).
Gore, R.: "Neptune—Voyager's Last Picture Show," *Nat'l. Geographic*, 34 (August 1990).
Gurnett, D. A., et al.: "First Plasma Wave Observations at Neptune," *Science*, 1494 (December 15, 1989).
Hammel, H. B.: "Neptune Cloud Structure at Visibile Wavelengths," *Science*, 1165 (June 9, 1989).
Hammel, H. B., et al.: "Neptune's Wind Speeds Obtained by Tracking Clouds in Voyager Images," *Science*, 1307 (September 22, 1989).
Hansen, C. J., et al.: "Surface and Airborne Evidence for Plumes and Winds on Triton," *Science*, 421 (October 19, 1990).
Helfenstein, P., et al.: "Large Quasi-Circular Features Beneath Frost on Triton," *Science*, 824 (February 14, 1992).
Hillier, J., et al.: "Voyager Disk-Integrated Photometry of Triton," *Science*, 419 (October 19, 1990).
Hubbard, W. B., et al.: "Interior Structure of Neptune: Comparison with Uranus," *Science*, 648 (August 9, 1991).
Ingersoll, A. P., and K. A. Tryka: "Triton's Plumes: The Dust Devil Hypothesis," *Science*, 435 (October 19, 1990).
Kerr, R. A.: "Triton Steals Voyager's Last Show," *Science*, 928 (September 1, 1989).
Kerr, R. A.: "The Neptune System in Voyager's Afterglow," *Science*, 1450 (September 29, 1989).
Kerr, R. A.: "Neptune's Triton Spews a Plume," *Science*, 213 (October 13, 1989).
Kerr, R. A.: "A Passion for the Little Things Among the Planets," *Science*, 998 (November 24, 1989).
Kerr, R. A.: "A Geologically Young Triton After All?" *Science*, 1563 (December 22, 1989).
Kerr, R. A.: "Geysers or Dust Devils on Triton?" *Science*, 377 (October 19, 1990).
Kinoshita, J.: "Neptune," *Sci. Amer.*, 82 (November 1989).
Kirk, R. L., Brown, R. H., and L. A. Soderblom: "Subsurface Energy Storage and Transport for Solar-Powered Geysers on Triton," *Science*, 424 (October 19, 1990).
Krimigis, S. M., et al.: "Hot Plasma and Energetic Particles in Neptune's Atmosphere," *Science*, 1483 (December 15, 1989).
Lane, A. L., et al.: "Photometry from Voyager 2: Initial Results from the Neptunian Atmosphere, Satellites, and Rings," *Science*, 1450 (December 15, 1989).
Lunine, J. I.: "Voyager at Triton," *Science*, 386 (October 19, 1990).
Lyons, J. R., Yung, Y. L., and M. Allen: "Solar Control of the Upper Atmosphere of Triton," *Science*, 204 (April 10, 1992).
McElheny, V.: "Neptune's Magnetic Looks," *Technology Review (MIT)*, 12 (January 1990).
Nelson, R. M., et al.: "Temperature and Thermal Emissivity of the Surface of Neptune's Satellite Triton," *Science*, 429 (October 19, 1990).
Ness, N. F., et al.: "Magnetic Fields at Neptune," *Science*, 1473 (December 15, 1989).
Pollack, J. B., Schwartz, J. M., and K. Rages: "Scattering in Triton's Atmosphere: Implications for the Seasonal Volatile Cycle," *Science*, 440 (October 19, 1990).
Polvani, L. M., et al.: "Simple Dynamical Models of Neptune's Great Dark Spot," *Science*, 1393 (September 21, 1990).
Poroco, C. C.: "An Explanation for Neptune's Ring Arcs," *Science*, 995 (August 30, 1991).
Smith, B. A., et al.: "Voyager 2 at Neptune: Imaging Science Results," *Science*, 1433 (December 15, 1989).
Smith, B. A.: "Voyage of the Century," *Nat'l. Geographic*, 34 (August 1990).
Soderblom, L. A., et al.: "Triton's Geyser-Like Plumes: Discovery and Basic Characteristics," *Science*, 410 (October 19, 1990).
Stone, E. C., et al.: "Energetic Charged Particles in the Magnetosphere of Neptune," *Science*, 1489 (December 15, 1989).
Stone, E. C., and E. D. Miner: "The Voyager 2 Encounter with the Neptunian System," *Science*, 1417 (December 15, 1989).
Strom, R. G., Croft, S. K., and J. M. Boyce: "The Impact Cratering Record on Triton," *Science*, 437 (October 19, 1990).
Stromovsky, L. A.: "Latitudinal and Longitudinal Oscillations of Cloud Features on Neptune," *Science*, 684 (November 1, 1991).
Suomi, V. E., Limaye, S. S., and D. R. Johnson: "High Winds of Neptune: A Possible Mechanism," *Science*, 929 (February 22, 1991).
Thompson, W. R., and C. Sagan: "Color and Chemistry on Triton," *Science*, 415 (October 19, 1990).
Tyler, G. L., et al.: "Voyager Radio Science Observations of Neptune and Triton," *Science*, 1466 (December 15, 1989).
Warwick, J. W., et al.: "Voyager Planetary Radio Astronomy at Neptune," *Science*, 1498 (December 15, 1989).
Yelle, R. V.: "The Effect of Surface Roughness on Triton's Volatile Distribution," *Science*, 1553 (March 20, 1992).

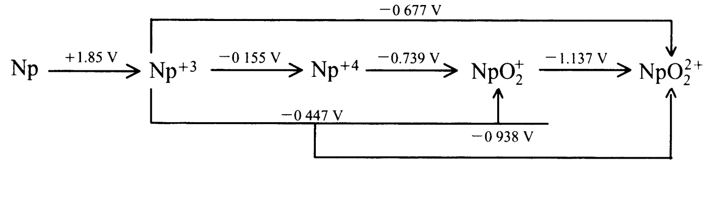
NEPTUNIAN SERIES. See **Radioactivity.**

NEPTUNIUM. Chemical element, symbol Np, at. no. 93, at. wt. 237.0482 (predominant isotope), radioactive metal of the Actinide series, also one of the Transuranium elements. Neptunium was the first of the Transuranium elements to be discovered and was first produced by McMillan and Abelson (1940) at the University of California at Berkeley. This was accomplished by bombarding uranium with neutrons. Neptunium is produced as a by-product from nuclear reactors. ^{237}Np is the most stable isotope, with a half-life of 2.20×10^6 years. The only other very long-lived isotope is that of mass number 236, with a half-life of 5×10^3 years.

^{237}Np is parent of the neptunium $(2n + 1)$ alpha decay series. Other isotopes include those of mass numbers 229–235 and 238–241; metastable forms of ^{236}Np, ^{240}Np and two of ^{237}Np are known. Electronic configuration $1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^2 4p^6 4d^{10} 4f^{14} 5s^2 5p 5d^{10} 5f^6 6s^2 6p^6 6d^1 7s^2$. Ionic radii Np^{4+} 0.88Å; Np^{3+} 1.02Å (Zachariasen). Oxidation potential $Np \rightarrow Np^{3+} + 3e^-$, 1.85 V; $Np^{3+} \rightarrow Np^{4+} + e^-$, -0.155 V; $Np^{4+} + 2H_2O \rightarrow NpO_2^+ + 4H^+ + e^-$, -0.739 V; $NpO_2^+ \rightarrow NpO_2^{2+} + e^-$, -1.137 V. See also **Chemical Elements.**

Neptunium has the oxidation states (VI), (V), (IV), and (III) with a general shift in stability toward the lower oxidation states as compared to uranium. The compounds which are formed are very similar to the corresponding compounds of uranium.

The ionic species corresponding to the oxidation states vary with the acidity of the solution; in acid solution of moderate strength the species are Np^{3+}, Np^{4+}, NpO_2^+, and NpO_2^{2+} as in the case of uranium and plutonium. The potential scheme in $1 - M$ HCI is as follows:



It will be seen that the metal is highly electropositive, in common with the other actinide elements. The $Np^{3+} \rightarrow Np^{4+}$ couple is reversible and this oxidation can be accomplished by the oxygen of the air. The (IV) state is stable, not oxidized by air, and only slowly oxidized to NpO_2^+ by nitric acid. The $Np^{4+} \rightarrow NpO_2^+$ couple is not readily reversible, whereas the $NpO_2^+ \rightarrow NpO_2^{2+}$ couple is reversible; this is reasonable on the basis that the former involves making or breaking the neptunium-oxygen bonds, whereas the latter does not. The oxidation of NpO_2^+ to NpO_2^{2+} requires moderately strong oxidizing agents. Neptunium differs from uranium and plutonium in that its potential relations are such as to render NpO_2^+ moderately stable with respect to disproportionating, even in solutions containing moderate concentrations of hydrogen ion.

The potentials are altered extensively by change in the hydrogen ion concentration and by the presence of any of a number of anions capable of forming complex ions.

Neptunium ions in aqueous solution possess characteristic colors: pale purple for Np^{3+}, pale yellow-green for Np^{4+} green-blue for Np^{5+}, while NpO_2^{2+} varies from colorless to pink or yellow-green depending on the acid present.

The precipitation reactions of Np^{3+} are similar to those of the tripositive rare earths, those of Np^{4+}, to the other tetrapositive actinides and to Ce^{4+}, and those of NpO_2^{2+} to the corresponding ions of uranium and plutonium. All of the simple salts of NpO_2^+ appear to be soluble.

The neptunium oxide system exhibits complexity similar to that found in the uranium oxide system. Thus, the important oxide is NpO_2 and there exists a range of compositions, depending upon conditions, up to Np_3O_8.

As a metal, Np has a relatively low melting point (~640°C), is very dense (20.45 g/cm^3), and is ductile. The alpha form reacts with hydrogen, carbon, oxygen, sulfur, the halogens, and phosphorus to yield a number of binary compounds.

The important halides of neptunium are the trifluoride, NpF_3, purple or black and hexagonal, the hexafluoride, NpF_6, brown and orthorhombic, the trichloride, $NpCl_3$, white and hexagonal, the tetrachloride, $NpCl_4$, red-brown and tetragonal, and the tribromide, $NpBr_3$, α-form green and hexagonal, β-form green and orthorhombic.

In research at the Institute of Radiochemistry, Karlsruhe, West Germany, during the early 1970s, investigators prepared alloys of neptunium with iridium, palladium, platinum, and rhodium. These alloys were prepared by hydrogen reduction of the neptunium oxide in the presence of finely divided noble metals. The reaction is called a *coupled reaction* because the reduction of the metal oxide can be done only in the presence of noble metals. The hydrogen must be extremely pure, with an oxygen content of less than 10^{-25} torr.

Industrial utilization of neptunium has been very limited. The isotope ^{237}Np has been used as a component in neutron detection instruments. Neptunium is present in significant quantities in spent nuclear reactor fuel and poses a threat to the environment. A group of scientists at the U.S. Geological Survey (Denver, Colorado) has studied the chemical speciation of neptunium (and americium) in ground waters associated with rock types that have been proposed as possible hosts for nuclear waste repositories. See Cleveland reference.

Additional Reading

Cleveland, J. M., Nash, K. L., and T. F. Rees: "Neptunium and Americium Speciation in Selected Basalt, Granite, Shale, and Tuff Ground Waters," *Science*, **221**, 271–273 (1983).

Keller, C., and B. Erdmann: "Preparation and Properties of Transuranium Element–Noble Metal Alloy Phases," *Proc. 1972 Moscow Symp. Chem. Transuranium Elements*, 1976.

Krot, N. N., and A. D. Gel'man: "Preparation of Neptunium and Plutonium in the Heptavalent State," *Dokl. Chem.* **177**, 1–3, 987–989 (1967).

Lewis, R. J., Sr.: "Hawley's Condensed Chemical Dictionary," Van Nostrand Reinhold, New York, 1992.

Magnusson, L. B., and T. J. LaChapelle: "The First Isolation of Element 93 in Pure Compounds and a Determination of the Half-life of $_{93}Np^{237}$," *Amer. Chem. Soc. J.*, **70**, 3534–3538 (1948).

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

McMillan, E., and P. H. Abelson: "Radioactive Element 93," *Phys. Rev.*, **57**, 1185–1186 (1940).

Seaborg, G. T.: The Chemical and Radioactive Properties of the Heavy Elements," *Chem. Engng . News*, **23**, 2190–2193 (1945).

Seaborg, G. T., and A. C. Wahl: "The Chemical Properties of Elements 94 and 93," *Amer. Chem. Soc. J.*, **70**, 1128–1134 (1948).

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

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

NERNST EFFECT. If heat is flowing through a strip of metal and the strip is placed in a magnetic field perpendicular to its plane, a difference of electric potential develops between the opposite edges. This phenomenon, discovered by Nernst in 1886, is analogous to the Hall effect, but with a longitudinal flow of heat replacing the longitudinal electric current. See also **Hall Effect.**

NERNST HEAT THEOREM. For a homogeneous system, the rate of change of the free energy with temperature, as well as the rate of change of heat content with temperature, approaches zero as the temperature approaches absolute zero.

NERNST-THOMPSON RULE. A solvent of high dielectric constant favors dissociation by reducing the electrostatic attraction between positive and negative ions, and conversely a solvent of low dielectric constant has small dissociating influence on an electrolyte.

NERVOUS SYSTEM AND THE BRAIN. As the end of the 20th century approaches, the neurosciences are intensely active, but they have not achieved a semblance of maturity. Although some knowledge of the human nervous system dates back nearly two centuries, numerous uncertainties pertaining to the fundamental principles of neuroscience remain to be clarified. Just within the relatively past few years have molecular and genetic science, as examples, been applied to this multifaceted and very intricate field. A flood of recent literature and the lively interchanges of viewpoints among the experts are indicative of a science that still is searching for agreement on various fundamentals. Only in the recent past have techniques, such as *patch-clamp* experimentation, been available to increase the resolution of research measurements. Progress is being made to expand and to meld the many disciplines that participate in the area of neuroscience. Unifying principles for the field remain to be established.

Although not unique in its awareness, the U.S. Congress designated the 1990s as the "Decade of the Brain." But it is interesting to note that the National Institute of Neurological Disorders and Stroke, which required over $700 million for the first year, of the decade, was awarded a smaller increase of less than $500 million.

Organization for Research. The manner in which neuroscience programs are organized depends, of course, on specific interests and whether these be private or governmental. The National Institute for Mental Health (U.S. Public Health Service) has several research branches:

- Behavioral and cognitive sciences
- Prevention and behavioral medicine
- Neuroimaging and applied neuroscience
- Personality and social processes
- Psychopharmacology
- Molecular and cellular neuroscience

The categorization of the neurosciences has been the subject of debate that reflects similar debates in other fields of scientific endeavor. On the one hand, there is a trend to create macrosciences made up of numerous micro- or minisciences, a situation that has prevailed in the areas of physics and chemistry. Conversely, there is a trend to create classifications that unite all sciences under one umbrella for a given topical field. Thus, several years ago, one commenced to hear of neuroscience used as a unifying term and that continues to persist. Thus created are such areas as neuroanatomy, neurochemistry, neuropharmacology, neurology, et al. A relatively recent term used is neurocomputing, and the list of specialties is expanding. In a recent countertrend, some of the neurosciences are being treated on a less specialized basis, such as placing cellular neurobiology in with other cell biology programs, and including developmental neurobiology in with developmental biology. Some researchers prefer more a specialized grouping of the neurosciences based upon the observation, "The brain has unique problems that are not shared by other systems."

Research Highlights. Among major accomplishments in the neurosciences since the mid-1980s include: (1) The discovery of scores of neurotransmitters (only four were known in the 1970s). (2) Learning that the brain, among its many other purposes, also functions as an endocrine gland by synthesizing and releasing several hormonelike peptides. It has been qualitatively established that these substances affect behavior, appetite, reproductive activity, and alleviate pain under certain circumstances. (3) Treatment of brain and nervous system illnesses and disorders, including Parkinson's disease, myasthenia gravis, Huntington's disease, depression, schizophrenia, and Alzheimer's disease, among several others. Advances have occurred, not because there is a full understanding of causation, but through the use of qualitative and clinical observations and experiments. (4) Initial studies that are attempting to relate brain and nervous system function with specific genes, such investigations assisted by probing involved chemicals and their reactions at the molecular level. (5) Neurosurgeons have dealt successfully with nerve transplantation and brain grafts. (6) Considerable research has been directed to describing and isolating specific nerve growth factors and relating these findings to general developmental physiology. (7) An encouraging start has been made in computerized brain mapping. (8) Laboratory research tools and instrumentation have undergone gross improvements, including the development of the patch-clamp technique and refinements of imaging methods. (9) Reviews of the literature reveal much increased activity in developing postulations and theories of neural functions at the neuron and total system level.

Several of the foregoing topics, particularly pertaining to nervous system disease and mental illness and to the function of the human visual and audio perceptive systems, are described elsewhere in this encyclopedia. Consult alphabetical index.

The Nervous System Complex

Like any other tissue of the body, the brain and nervous system are made up of cells. These cells function in accordance with the same fundamentals that apply to other cells, even though the cells of the complex interwoven brain tissue and the highly specialized cells of the central and peripheral nervous systems engage in unique operations. On a comparatively large scale, investigations of the brain and of the central nervous system can be conducted much as other organs are studied. Maps have been drawn of the interconnections of the brain on a relatively massive, regional basis, but remain severely lacking in terms of the huge numbers (millions or even billions) of interconnections among extremely tiny elements that constitute the infrastructure of the brain. Both the neuroanatomy of the neurophysiology of the brain, can be described as in a very early stage of understanding. Most scientists will agree that this statement holds even in view of the very impressive progress that has been made in brain and nervous system research over the last several years. Up to a point, the electrical and chemical signalling apparatus of the brain and nervous system can be measured and interpreted, but beyond a general understanding of the brain, the problem of comprehending brain function in depth is of a staggering nature—and for several reasons.

(1) *The sheer numbers* of individual elements of the brain system that exist within an organ whose mass in the average adult human is about 45 ounces (1.3 kilograms), representing an information processing system unparalleled in the universe as we know it today. It is conservatively estimated that there are on the order of 10^{11} (a hundred billion) nerve cells (neurons) in the human brain. This number approximates that of the number of stars in our galaxy. Some neuroscientists consider this figure woefully small. If the 10^{11} figure is assumed, it suggests that neurons in the embryonic and fetal stages, originating as a flat sheet of cells on the dorsal surface (*neural plate*) of the developing embryo, must develop at an average rate of more than 250,000 per minute. Once the production of neurons is completed, there is no further replication during the remainder of life, but as life progresses, large numbers of neurons are destroyed.

How can researchers "get a handle" on so many components? By comparison, the number of components in the most capable of modern computers pales. The number of synapses (*connections*) in the brain is estimated at 10^{14} (100 trillion).

(2) *The extremely small size* of the nerve cells is another problem. Research has confirmed that so many "components" are perforce extremely tiny to fit in such a relatively small volume. The typical neuron (cell) is estimated to range from 5 to 100 micrometers (thousandths of a millimeter) in diameter. There are even much tinier subelements which make up these cells. Very important nerve actions occur in which the amount of substance involved is expressed in terms of a few hundreds of molecules.

(3) *The variations of form and function of neurons* make generalizations difficult. The neuron is not a conventionally shaped cell, but takes the form of a trunk of a tree with a well-developed root system, where the trunk may be as short as 10–12 micrometers (thousandths of a millimeter) in diameter and yet the entire cell may have a length ranging from as short as 0.1 millimeter to as long as 1 meter or more. Because neurons are tightly packed in many regions of the brain, they tend to occur in "thickets," much as the intertwining roots of closely-planted trees. But, despite the high-density packing, a very effective insulation between the intertwining fibers provides full chemical and electrical integrity of each neuron except at specific designated points (*synapses*) of connection. The fluid film which provides this protection of nerve fibers is only about 0.02 micrometer in thickness.

As will be evident from later descriptions, there are many variations in the physical and electrical and chemical characteristics among individual neurons and, to date, these variations have not been easy to classify in terms of a relatively few categories. An important aspect of progress in neuroscience during recent years has been that of identifying previously unknown neuron forms and characteristics and developing a series of generalizations that provide pathways toward a better understanding of specific information. Alrhough from a systems standpoint, there appears to be a high degree of specificity represented by the neurons and their synapses, the tight packing of these elements into such a small volume (skull) creates a type of geometry that tends to defy the neatness desired by the human analytical mind. It is as though millions of very neatly arranged computer circuit boards or silicon chips were tightly squeezed into a plastic ball, but with the integrity of these elements still maintained. Although the analogy is not very apt, the intricacies of the brain resemble the "spaghetti" of wiring encountered in an old-fashioned radio chassis as contrasted with the extreme neatness and geometry of modern electronic equipment.

(4) *The dynamics of the brain system* can only be partially studied by examining dead brain tissue. It will be recalled that death itself is defined in terms of "brain death." It is the living, functioning brain that yields a full comprehension of brain function. Lower animals and some primates have provided leads to understanding the human brain function. Much of what has been learned concerning the nerve impulse has been gleaned from studies of the squid.

Research Breakthroughs of the Past. Because of the very high-density packing of neurons in brain tissue, a discovery by the Italian anatomist Camillo Golgi, as early as 1875, was of great note. Although the phenomenon of Golgi's brain tissue staining technique remains unexplained, the technique makes it possible to stain only small numbers of brain cells at any one time. This technique has proved invaluable to neuroanatomy and neurophysiology because the method allows specific identification and study of a few components that otherwise would be lost in a morass of thousands upon thousands of similar components. A Spanish contemporary of Golgi, Ramón y Cajal, applied Golgi's technique over a lifetime and in 1904 published "*Histologie du système nerveux de l'homme et des vertébrés.*" This publication is still regarded as the greatest single work in neurobiology. In the time of Golgi and Cajal, the research tools were confined to the staining technique and the light microscope. One of Cajal's sketches of Golgi-stained nerve tissue is shown in Fig. 1. Within these limitations, Cajal proposed two important concepts: (1) Patterns do not suggest a network (particularly as suggested by a modern computer network), but rather a system comprised of specific cells which communicate with each other only at certain points (connection points or synapses). (2) Interconnections do not suggest a random arrangement, but rather a highly structured, quite specific arrangement. One might suggest that perhaps the organization of the brain may provide new concepts in computer engineering that can be reduced to practice only after a great deal more is learned of brain function and organization. Based upon present knowdedge, the often suggested brain-computer analogy is of little value.

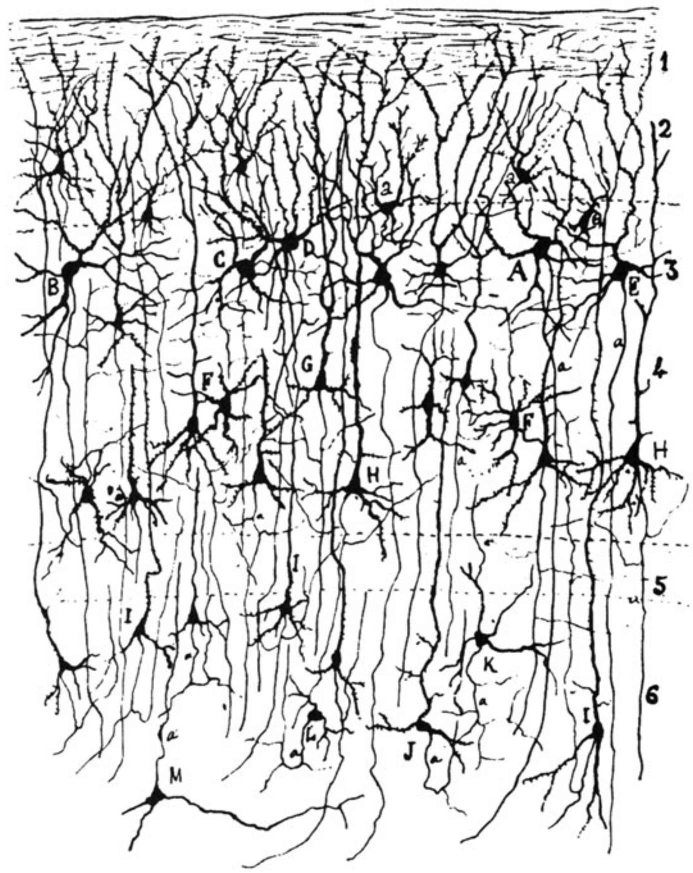


Fig. 1. Sketch from "Histologie du système nerveux de l'homme et des vertébrés," published by Ramón y Cajal in Spain in 1904. Cellular layers are identified by numerals at right-hand edge of diagram. Neurons are indicated by capital letters.

It was not until the early 1950s that a scientist in the Netherlands. Walle J. H. Nauta, developed an improved staining method. Nauta found that there is a short period of the time during which the nerve fiber degenerates when the fiber can be stained to differentiate it from living fibers. Thus, it is possible to trace that fiber to a specific neuron that may be remotely located in some other region of the brain. This technique has greatly expanded the researcher's ability to trace interconnections and thus add many details to the maps of the brain.

Since the mid-1960s, other advanced techniques have contributed in a major way to developing a better understanding of the anatomy of the brain. One of these is the injection of radioactive substances into brain tissue. Some of these substances are absorbed by cell bodies and transported to terminals; other substances are absorbed by the terminals and transported by axons to the cells. In a technique developed by Louis Sokoloff (National Institute of Mental Health), radioactively labeled deoxyglucose is used. This substance collects in nerve cells and the degree of accumulation is an indicator of a cell's activity. This technique is used in conjunction with known laboratory stimulation of various areas of the brain and thus permits brain research in real time as contrasted with examining dead tissue. For example, if the stimulation ia audio, those cells which are involved in the process of hearing will be active and can be revealed by later microscopic examination. In a more recently developed technique, positron-emitting radioactive isotopes are combined with tomography which makes it possible to conduct these studies entirely outside the skull and thus makes it possible to expand the reseach to living animals and, in the diagnosis of mental problems, the technique may be refined to include observations of the living human brain.

The decades of the 1970s and 1980s were marked by numerous discoveries pertaining to the biochemistry of the brain. For example, the known chemical transmitters has been expanded from a list of just a few to several. Much has been learned concerning the production of chemical substances within the brain, once considered a function in the exclusive province of the glands. Research commenced by neuroscientists at McGill University in the mid-1950s led to the discovery of "pleasure centers" within the brain that play a role in learning and memory. As pointed out by Routtenberg (1978), the so-called reward system can be localized in particular nerve cells and their fibers. The system is affected by drugs that interact with the substances secreted by these nerve cells. The fact that only certain nerve-fiber pathways are implicated in brain reward suggests that these pathways have a specific function. This parallels what is known of the brain's visual and movement systems, each of which has a specified set of component pathways. This research led to many new insights on the functioning of various psychoactive drugs. Neuroscientists at several institutions have found many new substances, generally referred to as neuropeptides, in brain tissue. These substances include the heavily publicized enkephalins and endorphins. Researchers have found that neuropeptides differ from previously identified transmitters in that they appear to orchestrate complex phenomena, such as thirst, memory, and sexual behavior. See also Table 1.

Broad, Traditional Concepts of the Brain

The large, soft mass of nerve tissue making up the brain is contained within the cranium (*encephalon*). The brain consists of four major parts: (1) the *cerebrum*; (2) the *cerebellum*; (3) the *pons Varolli*; and (4) the *medulla oblongata*. The brain and spinal cord together constitute the *central nervous system*.

By way of generalization, the brain is the control station for nerve impulses. The brain is composed chiefly of nerve cells with their fibers interwoven in a complex relay system. At birth, the human brain weighs between 11 and 13 ounces (312 and 369 grams) and, as previously mentioned, attains a weight of about 45 ounces (1.3 kilogram) in the adult. The brain increases in weight until about the twentieth year, after which there is gradual loss of weight for the remainder of life. The volume of the human adult brain is 2000 cubic centimeters or greater. By comparison with other animals, the human brain is very large. See Fig. 2.

In 1990, paleoanthropologist Dean Falk postulated that the hominid brain required a special cooling system before the brain could expand to its present size. The scientist proposed that the brain size, like an engine, is restricted by the capacity of the cooling system if it is to continue "running" within a given temperature range. Simply called the "radiator theory," Falk's proposal has garnered both positive and negative reactions from professionals in the field. Pro and con details are covered by A. Gibbons in the listed reference.

In an excellent paper, P. H. Harvey and J. R. Krebs (University of Oxford, Department of Zoology) point out, "Some animals have larger brains than others, but it is not yet known why. Specific differences in life-style, including dietary habits and patterns of development of the young, are associated with variation in brain weight, independently of the effects of body weight and evolutionary history. Taken together with behavioral and neuroanatomical analyses, these studies begin to suggest the evolutionary pressures that favor different sized brains and brain components." In their paper, the investigators compare brain size and functionality against numerous other body features in numerous animals as well as in humans. It is pointed out that it has not been unusual for researchers to seek associations between encephalization and more than 20 variables that summarize the life-styles of various species. Among specific animal groups, a significant association at the 5% probability level includes: (1) diet in mammals and (2) the pattern of development in birds.

In any investigation of encephalization, it is evident that, when one brain region is enlarged, the whole brain is not. The authors conclude their paper with an observation of Mark Twain: "I never could keep a promise. I do not blame myself for this weakness, because the fault must lie in my physical organization. It is likely that such a liberal amount of space was given to the organ which enables me to make promises that the organ which should enable me to keep them was crowded out."

Medulla Oblongata. This organ is approximately $\frac{3}{4}$ of an inch to 1 inch (19 to 25 millimeters) in length. Externally, it appears like an expanded part of the spinal cord. See Figs. 3 and 4. Internally, its structure is quite complex and consists of nerve tracts passing into the brain. From some of the nuclei come fibers that eventually emerge to form the VIIIth, IXth, Xth, XIth, and XIIth cranial nerves. Cell centers in this area also are concerned with swallowing, vomiting, breathing, speech,

TABLE 1. VERY ABRIDGED LIST OF IMPORTANT NERVOUS SYSTEM AND BRAIN RESEARCH EVENTS

- In 1870, G. T. Fritsch and E. Hitzig first suggested functional compartmentalization in the cerebral cortex.
- In 1872, Francisci Gennari (Parma, Italy) discovered the visual cortex.
- In 1875, Golgi developed a brain tissue staining technique which enables the staining of a comparatively few neurons and neuronal systems at one time, thus making it possible to study a few out of many thousands of adjacent systems.
- In 1885, P. Ehrlich discovered the blood-brain barrier, finding that many blood-borne solutes do not penetrate into central nervous system tissue as rapidly as they penetrate into most other tissues. The functional significance of the blood-brain barrier mechanism is to buffer the neuronal microenvironment against changes in plasma concentrations of various important solutes and to regulate the composition of the neuronal "atmosphere" for optimum performance. See **Blood-Brain Barrier.**
- In 1904, Cajal published the findings of a lifetime of study of brain tissue and neuronal systems as the result of combining the Golgi staining technique with light microscopy.
- In 1904, Ivan Pavlov, Russian physiologist, was awarded the Nobel prize in medicine for pioneering studies of the nervous system and demonstrating the conditioned reflex in dogs.
- In the first third of the present century, numerous investigators, among them Sir Henry Dale, Otto Loewi, A. L. Hodgkin, A. F. Huxley, B. Katz, Sir John Eccles, and S. W. Kuffler, made major contributions to an understanding of how individual neurons work. One finding was that all neurons, regardless of size and shape, appear to utilize the same two kinds of electric signals (graded potentials and action potentials). Dale initially proposed the concept that a neuron released only one transmitter chemical from all its terminals, an observation that was considered inviolable until recently. It is now suspected that some neuropeptides can coexist in the same neurons as certain transmitters, such as norepinephrine and serotonin.
- In 1949, Donald O. Hebb published "The Organization of Behavior," which became a keystone of modern neuroscience. The tome described cell assembly theory and foreshadowed neural network theory.
- In the 1950s, E. J. Furshpan and D. D. Potter (University College London) discovered that some synapses are profoundly different from the usual chemical type, depending on the flow of current rather than diffusion of a transmitter.
- In the early 1950s, W. J. H. Nauta (then in the Netherlands) developed an improved brain tissue staining technique which differentiated degraded nerve fibers from living fibers, making it possible to identify specific fibers with remotely located neurons. This expanded the ability of researchers to provide greater detail in brain mapping.
- In the early 1950s, A. L. Hodgkin, A. F. Huxley, and B. Katz (British scientists), as the result of studying the nerve-impulse transmission in the giant axon of the squid, demonstrated that the propagation of the nerve impulse coincides with sudden changes in the permeability of the axon membrane to sodium and potassium ions.
- In the early 1950s, P. A. Weiss and colleagues (University of Chicago) discovered the phenomenon of axonal transport.
- In the mid-1950s, Sutherland and his associates (Case Western Reserve University) demonstrated that dopamine and norepinephrine, among other transmitters, increase or decrease the concentration of a second messenger substance in target cells, the latter mediating the electric or biochemical effects of the transmitter in the first messenger. (See later reference to Sutherland.)
- In the mid-1950s, neuroscientists at McGill University (Montreal) discovered pleasure centers within the brain that play a role in learning and memory.
- In the 1960s, methods for selectively staining neurons containing a particular transmitter were developed and utilized by a number of investigators. The natural transmitter substance can be converted into a fluorescent derivative that glows under ultraviolet radiation. Another method was developed that takes advantage of the high specificity of antibodies. A specific enzyme involved in the synthesis of a particular transmitter is purified from brain tissue and then injected into an experimental animal, whereupon the enzyme induces the manufacture of antibodies which, in turn, specifically combine with the enzyme. The antibodies are then purified and labeled with a fluorescent dye. Then they can be used to selectively stain neurons containing the relevant enzyme. Such techniques have greatly expanded the researcher's ability to map detailed anatomical distribution of individual transmitters. Falck (University of Lund) and Hillarp (Karolinska Institute, Sweden) first demonstrated that neurons containing monoamines fluoresce green or yellow when the transmitters are first converted into fluorescent derivatives. The best mapped transmitters are the monoamines norepinephrine, dopamine, and serotonin.
- In the early 1960s, V. P. Whittaker (University of Cambridge) and E. De Robertis (University of Buenos Aires) found that when brain tissue is gently disrupted when homogenized in a sugar solution, many nerve terminals break away from their axons. These terminals form intact, closed particles called *synaptosomes* which contain the means for synthesis, storage, release, and transmitter inactivation associated with the nerve terminal. The synaptosomes can be purified by spinning in a centrifuge. The availability of these substances made it possible for neuroscientists to study the mechanisms of synaptic transmission in vitro.
- In 1962, F. H. C. Crick was awarded the Nobel prize in physiology or medicine for contributions to the understanding of how the human brain works.
- In the 1970s, L. L. Iversen (Cambridge), T. J. Crow (London), and P. Seeman (University of Toronto), among others, revealed abnormally high concentrations of dopamine and dopamine receptors in the brains of deceased schizophrenics.
- In the 1970s, Perry (University of British Columbia) discovered that a specific deficit of gamma-aminobutyric acid (GABA) occurs in the inherited neurological syndrome known as Hungtington's chorea. See **Chorea (Huntington's).**
- In the 1970s, pioneering research efforts revealed that opiate receptors are located in those regions of the brain and spinal cord which are known to be associated with pain. S. H. Snyder and C. B. Pert (Johns Hopkins University School of Medicine), E. J. Simon (New York University School of Medicine), and L. Terenius (University of Uppsala), among others, used radioactively labeled opiate compounds to reveal these sites. This research paved the way for discovery of the enkephalins and endorphins, among other neuropeptides.
- In 1971, Bloom and co-workers (National Institute of Mental Health) demonstrated that cyclic AMP can affect signaling in neurons. A bit later, Greengard and associates (Yale University School of Medicine) showed that cyclic AMP is involved in the synaptic actions of several brain transmitters (norepineprhine, dopamine, serotonin, histamine). Greengard et al. suggested a unifying process to the effect that cyclic AMP activates specific enzymes in the target cell called *protein kinases* and, through a complex process, changes the level of excitability of the target cell.
- In 1971, Sutherland (see previous mention in this list) received the Nobel prize in physiology and medicine for identifying the second messenger substance as cyclic adenosine monophosphate (cyclic **AMP**).
- In the early 1970s, L. E. Eng and A. Bignami (Stanford University) isolated glial fibrillay acidic protein (GFAP) as a component of astrocytes.
- In 1975, J. Hughes and H. W. Kosterlitz (University of Aberdeen) first isolated the enkephalins. Shortly thereafter, the endorphins, also morphine-like compounds, were isolated from the pituitary gland.
- In the late 1970s, researchers found "substance P," associated with spinal neurons involved in pain stimuli.
- In 1979, Rita Levi-Montalcini and Pietro Calissano were awarded the Nobel prize for physiology or medicine for their identification of the nerve-growth factor (NGF).
- In the early 1980s, research on trophic substances, such as nerve-growth factor (NGF), accelerated into a major effort.
- In 1991, Bert Sakmann and Erwin Neher shared the Nobel prize in physiology or medicine for their invention of the patch-clamp technique used in cell membrane and ion channel research.

digestion, metabolism, and the beating of the heart. In the medulla oblongata, the large bundles of fibers, which originated in the two halves of the cerebrum and which transmit the impulses of voluntary movement, cross to the opposite sides. Thus, movement in the right arm, for example, is controlled by centers in the left half of the cerebrum.

Pons. Lying above the medulla oblongata and continuous with it is the pons. It is made up of massive bundles of fibers that start in the cerebrum and sweep backward to the cerebellum. This connection makes possible many skilled acts that require coordination of sight, hearing, muscular movement, and various other sensations. The playing of a musical instrument is an example of such an act. The pons contains a space called the *fourth ventricle.* In the floor of this ventricle is the nucleus of the VIth cranial nerve. This nerve, which has the longest course inside the skull of all the cranial nerves, is concerned with turning the eyeball outward.

Cerebellum. The second largest part of the brain is in back of the pons and termed the cerebellum. It lies in the back of the skull. The cerebellum is made up of many narrow, leaflike folds arranged into two

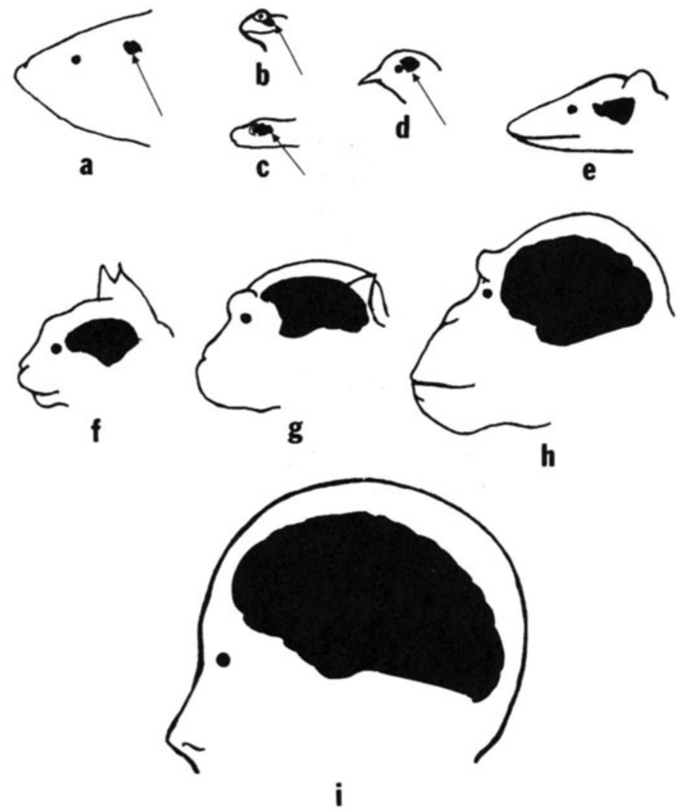


Fig. 2. Comparative sizes of the cerebrums of various animals: (a) bass, (b) leopard frog; (c) grass snake; (d) pigeon; (e) opossum; (f) cat; (g) macaque monkey; (h) chimpanzee; (i) human. (*After Hubel.*)

large masses, and of a middle portion. Rich in cells, it has many complex connections with the brain above it and with the spinal cord below. The chief function of the cerebellum is to coordinate relatively complex movements into special acts. This may be movement in different parts of the same limb; combined action of the head, limbs, and body. For example, picking up a pencil, writing with it, and laying it down again requires smooth interaction of many muscle groups. The cerebellum correlates the actions of the various groups. To do this, range, direction, rate, and force of movement must be synchronized and maintained with the movement of the eye. Disease in the cerebellum does not cause paralysis, but it does produce disturbance of muscular coordination. Tremors, staggering gait, and excessive relaxation of the muscles result from disease in this part of the brain, as well as disturbances in components of muscular activity.

Midbrain. This is a small area between the pons and the cerebrum. It is an important relay station for the sensory impulses. The midbrain also governs some muscle activity of a reflex nature. Many of the involuntary acts of the eye, such as narrowing of the pupil in bright light, originate in this area. The IIIrd, IVth, and Vth cranial nerves originate from cell collections in this area of the brain.

Hypothalamus. Just above the midbrain, an important group of nuclei comprise the hypothalamus. Beneath it, the two large nerves from the eyes meet and part of their fibers cross to the opposite sides. Other cells of this region are concerned with such vital functions as regulation of body temperature, metabolism, and heart rate. Sexual development, sleep, and the body's use of fat and water are influenced by this region in the brain. There is an essential relationship between the hypothalamus and the nervous system and the endocrine system.

Thalamus. The thalamus is found next to the foregoing group of cells and contains another group of nuclei which integrate sensations of many sorts. Also, the thalamus is the site of a crude form of consciousness and plays a role in the production of emotion. When this part of the brain is diseased, spontaneous laughter or crying may occur. The

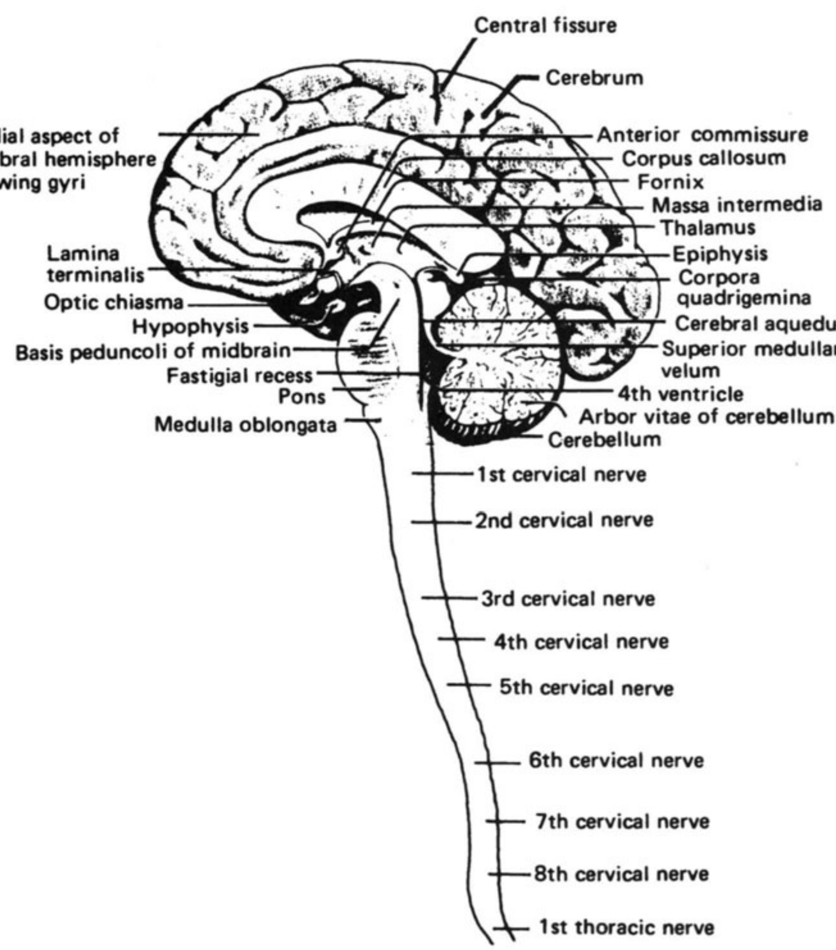


Fig. 3. Traditional diagram of median sagital section of the human brain.

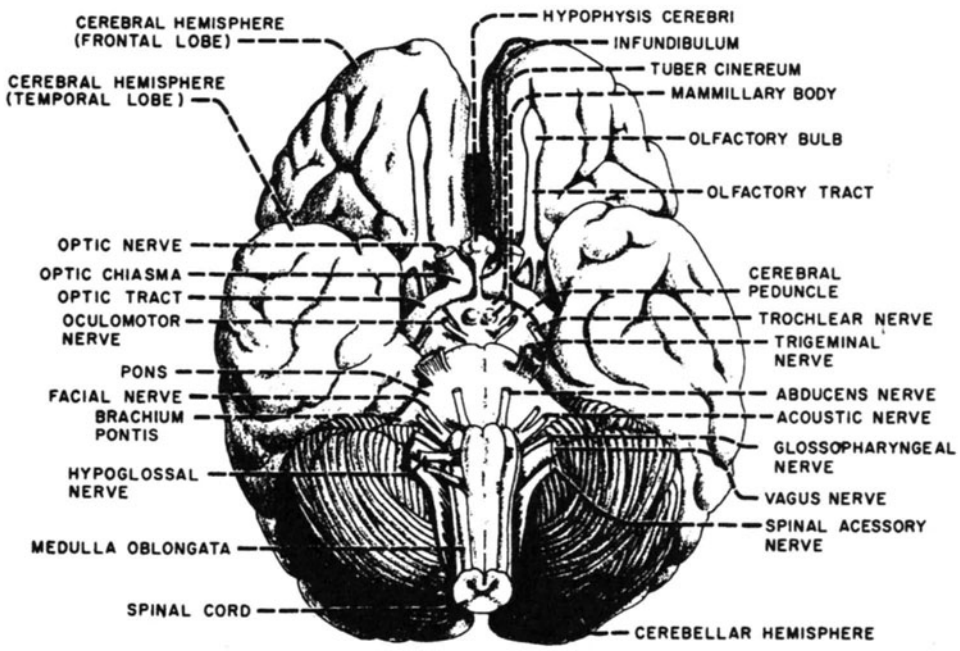


Fig. 4. Traditional diagram of the base of the human brain and cranial nerve roots.

crude emotional responses that arise are further elaborated and controlled by the cerebral cortex.

Hippocampus and Amygdala. Both of these structures are anchored to the inner surface of the temporal lobe in both hemispheres of the brain. Because of its shape, hippocampus is named from the Greek for *seahose*. Similarly, because of its shape, the amygdala is named from the Greek for *almond*. From experience with amnesia patients and experimentation with laboratory animals, it has been noted that memory loss occurs after surgery or damage to the hippocampus. Neuroanatomical studies of the circuitry linking the structures indicates similar, but not identical participation of these two structures. Mishkin reported in 1987 that although the hippocampus and the amygdala can substitute for each other in learning to recognize an object, the hippocampus seems to be particularly important for learning spatial relations. Further research has shown that the amygdala (amygdaloid complex), which consists of several nuclei, has direct and extensive connections with all the sensory systems in the cortex. The amygdala also communicates with the thalamus, along a path that is a segment of the memory system. Mishkin further observes that the same parts of the amygdala on which sensory inputs converge send fibers deeper into the brain to the hypothalamus (believed to be the source of emotional response). The amygdala, with so many connections, is now believed to mediate the association of memories formed through the different senses.

Contribution of the hippocampus to the memory system has been studied in monkeys by L. R. Squire and S. Zola-Morgan (Veteran Affairs Medical Center, San Diego, California) and reported in 1991. See reference listed. The paper concludes with, "Cumulative and systematic research with monkeys and related research with humans has identified the components of the medial temporal lobe memory system: the hippocampus, together with adjacent anatomically related cortex (entorhinal, perirhinal, and parahippocampal cortex). This system is fast, has limited capacity, and performs a crucial function at the time of learning in establishing long-term declarative memory. its role continues after learning during a lengthy period of reorganization and consolidation, whereby memories stored in the neocortex eventually become independent of the medial temporal lobe memory system. This process, by which the burden of long-term (permanent) memory storage is gradually assumed by the neocortex, assures that the medial temporal lobe system is always available for the acquisition of new information. The anatomy and function of the system and its relation to the neocortex are becoming well enough understood that computational modeling should provide a fruitful way to make these ideas more formal and quantitative."

Cerebral Hemispheres. The two cerebral hemispheres represent 70% of the entire nervous system. This is the area of the nervous system in which all the sensory experiences are mixed and blended. Specific sensory impulses thus become associated with many others and expand the experience and consciousness. The individual's capacity for many and varied activities, and memory, emotions, and ideas is dependent upon the action of this part of the nervous system.

The surfaces of the hemispheres are marked by large, rounded folds and deep grooves. Partly on the axis of the main grooves and partly on imaginary lines, the cerebrum is divided into four lobes: *frontal, parietal, temporal*, and *occipital*. Each lobe has special functions, but these functions are only partly understood.

The occipital lobes at the back of the skull are the site where visual impressions are made. Color, size, form, movement, and distance are evaluated in this portion of the brain, leading to the identification of a particular object. Also, the differences between similar objects are discerned. For example, two objects high in the air can be recognized as a bird and an airplane on the basis of past experience. Injury to this area may cause blindness.

The temporal lobes receive the fibers concerned with hearing, speech, balance, and smell. Diseases related to these lobes cause loss of smell, or they may be responsible for imaginary smells.

The parietal lobes are concerned with taste sensations and some other sensations, such as the ability to judge weight, shape, and textures. By the action of this area, one is able to tell what various objects are by feeling, rather than seeing them.

The frontal lobes are concerned with some of the most complex abilities of the mind. Reason, emotions, and judgment have their site here. Additionally, there is a group of large cells in the posterior region of the frontal lobes which are involved with complex voluntary movements. The speech center is located here. It is found to predominate on the left side in right-handed individuals and vice versa. If the speech function is lost because of injury to only one side of the brain, it can sometimes be reacquired by reeducation. The area responsible for these complex voluntary movements is called the *motor cortex*. The muscles of the body are controlled by various areas of the motor cortex. Irritation of the cells in these zones will cause spasms of the muscles they supply. Destruction of the cells will produce complete loss of voluntary movement of the muscles. These cells also function to keep the muscles in balance between relaxation and contraction. If this region of the brain is seriously damaged or destroyed, this inhibiting power is lost. Consequently, the muscles become contracted and stiff (*spastic paralysis*).

The frontal lobes have many connections with the thalamus as well as with the other lobes of the brain. In the frontal lobes, feelings or emotions are added to the other associations. The combination of feeling and knowing determines most voluntary action of the body. Thinking, reasoning, judgment, and imagination result as the sensory and emotional associations become more complex. Disease in the frontal lobes of the cerebrum causes personality changes, errors in judgment and insight, and poor emotional control.

Twelve pairs of nerves arise from the brain proper. The 1st is associated with the sense of smell; the IInd with sight: the IIIrd and IVth with the eye muscles which move the eyeball or the muscles of the pupil; and the Vth causes the muscles of the jaw to move. The VIth is concerned with the movement of the eye to the side. The VIIth carries impulses to all the muscles of the face. The VIIIth conducts impulses having to do with hearing and with balance; the IXth transmits taste sensations from the posterior third of the tongue, and other sensations from the throat and mucous membranes; and also aids in swallowing. The Xth is an exceedingly long nerve that extends down the neck and into the chest and abdomen; it is concerned with swallowing and talking; its action also slows the rate of the heart and regulates the movement of the stomach. It is a large part of the parasympathetic system in the upper part of the body, including the esophagus, stomach, intestines, liver, bronchi, lungs, heart, and blood vessels. The XIth supples some of the muscles that turn the head and some of those in the neck. The XIIth is responsible for the movement of the tongue.

Spinal Cord. The nerve tracts passing to and from the brain are contained in the spinal cord, which is continuous with the lower part of the brain. It is about 18 inches (46 centimeters) long and is rounded in shape. This nerve tract is larger in the regions which give rise to the nerves to the arms and legs, since these parts have many complex functions, thus requiring a large nerve supply. From the neck to the lowest parts of the vertebral column, 31 pairs of nerves emerge from the spinal cord. Each nerve is attached to the cord by two roots. Because the spinal cord is not as long as the vertebral column, the roots of the nerves must gradually increase in length before they can emerge from between the vertebrae. These longer nerve roots collect in a mass that fills the lower end of the vertebral canal. The structure resembles a horse's tail and is called the *cauda equina*.

A cross section of the spinal cord reveals a gray figure, roughly shaped like an "H," imposed on a white background. The nerve cells make up the gray matter, while the nerve bundles form the white matter. Bundles with specific functions occupy specific areas of the spinal cord. Thus, injury to the cord will result in certain abnormal reactions which will be evident upon neurological examination. The abnormal findings will suggest where the diseased part is located.

Impulses which arise in the brain and are concerned with voluntary muscular movements are received by specific cells in the spinal cord. They relay these impulses to the nerves which control the various muscles. These cells have connections with other cells in the nervous system that act together to bring about *reflex activity*. Reflexes control the position of the head so that it automatically assumes the normal position. Withdrawal or *flexion* reflexes pull limbs away from painful or disagreeable stimuli. *Extensor* reflexes straighten out the limbs and work with the flexor reflexes. Bladder and bowel actions result from reflex action over which there is some voluntary control. In injury to the spinal cord, the tracts allowing voluntary control may be interrupted, so that action is then reflex in origin.

The cells on the front and anterior side of the spinal cord connect with cells in the cerebellum to control the direction and precision of normal muscular movement. From still another part of the brain, connecting fibers come to these cells to bring about certain automatic, associated muscular movements, as in the instance of swinging of the arms as one walks. Another function of these cells is the maintenance of the proper amount of constant contraction or *tone* of the muscles. If the muscles are too contracted, they move too slowly and rigidly. If they are too relaxed, too much stimulation is needed to make them respond.

The cells along the side of the cord send out fibers that unite with others to form the *sympathetic chain*. These cells are concerned with the action of involuntary muscles in the intestines, arteries, and other internal structures. Various glands receive fibers from these cells. Cells on the back or posterior side of the spinal cord receive the sensations of touch, pain, vibration, temperature, pressure, and position. They then transmit these various sensations to other cells in the brain. Thus, impulses of many sorts travel down the paths in the spinal cord, while others enter it and travel upward to the brain. Still others enter and travel only part of the way up and set off the spinal reflexes.

The Nervous System

The peripheral nervous system is composed of the cranial nerves, the spinal nerves, and the autonomic nervous system. The autonomic system supplies nerves to most of the "automatic" organs of the body, i.e., the glands, heart, blood vessels, and the involuntary muscles in the internal organs. One part of the autonomic nervous system prepares the individual for emergency situations, shifting circulating blood to skeletal and heart muscles, increasing heart and lung function, dilating the pupils of the eyes, and moistening the skin with perspiration. The other division is concerned with conserving and restoring bodily resources. Thus, it protects the eyes by causing the pupils to constrict in bright light, and prevents the heart from overexerting itself. All of the digestive processes are promoted by this division. See **Autonomic Nervous System.**

Nerve Definition and Function. A nerve is a slender cord made up of nerve fibers (neurons). In the vertebrates the nerve is surrounded by loose connective tissue called the epineurium. Each small bundle of fibers within the epineurium is surrounded by a thin, compact perineurium, and from this layer thin septa, the endoneurium, run between the irregular groups of fibers within the bundle.

Nerves form the communicating paths between the central nervous system and the various parts of the body, as well as the connections between ganglia. They have been given special names according to their anatomical distribution, and a few functional properties have been indicated by descriptive terms. Thus sensory or afferent nerves carry impulses from sense organs to the central nervous system and motor or efferent nerves lead out to muscles and other effectors. Most nerves of the body contain fibers of each kind and so are mixed nerves. Nerves arising from the brain in the vertebrates are called cranial nerves and those connected with the spinal cord are spinal nerves. Typically, all of these main nerves of the vertebrate are supposed to be based on the form of the spinal nerves, which connect with the cord by two roots. The dorsal root bears a spinal ganglion and carries all the sensory fibers and the ventral root lacks a ganglion and is motor. All sensory nerves bear a ganglion or arise from a sensory layer such as the retina of the eye and no purely motor nerves have ganglia.

The motor components of these nerves grow out from cells in the central nervous system and the sensory components grow into the central system from ganglia or from sensory cells and also out from the ganglia toward the periphery of the body.

A nerve impulse is a progressive transfer of a condition of excitation along a nerve fiber, initiated by a stimulus acting upon a sensory organ, by a cell within the nervous system, or experimentally by direct stimulation of the nerve fiber. The nerve impulse involves measurable chemical changes and energy consumption accompanied by changes in electrical potential associated with membrane depolarization. The electrical changes have been used extensively in investigation of nerve physiology and of the brain. The passage of the nerve impulse over the synapse is accompanied by the action of acetylcholine in cases which have been studied. Terminal nerve fibers may form either acetylcholine or sympathin, the latter being closely related to adrenalin. The impulse activates some organ of the body or enters the nervous system, where it is relayed to other parts. These mechanisms are further described later.

The attainment of harmonious action of all component parts of the animal and the adjustment of its behavior to environmental conditions is by way of communication and regulation through the nervous system. Coordinative processes of this type are aided by the secretion and distribution of hormones. The latter process is termed *chemical coordination*. Coordination by way of the nervous system occurs, of course, with much greater speed than actual materials can be transported from one part of the body to the next. The nervous system is responsible for the immediate adjustment of the animal to fluctuating conditions, while chemical coordination is extensively involved in the maintenance of the normal organic processes which are a more uniform part of the animal's activity. Coordination, in general, is due to close interaction of the two types; neither is wholly independent of the other.

The process of coordination through the nervous system is based on the general sensitiveness of protoplasm known as irritability and on the property of conductivity by which some result of stimulation passes rapidly through the adjacent substance. All living cells are irritable to some extent. They respond to some stimulus with characteristic activity. Sensory organs possess this property to a high degree and in addition are specialized to receive a certain kind of stimulus. Nervous tissue is specialized for ready activation and for the rapid conduction of the impulse generated within it.

The organs which receive stimuli as a special function for the benefit of the animal, whether relatively simple cells or extremely complex organs like the human eye, are known as receptors. They have special nerve endings, often associated with other structures, and are connected by afferent or sensory nerves with other parts of the nervous system or, in very simple animals, with some organ capable of acting in response to the condition from which the stimulus arose. Organs of the latter category are muscles, glands, electric organs, light organs, and some pigment cells, and are known collectively as effectors. Usually one or more cells of the nervous system intervene between the sensory cell associated with the receptor and the motor cell which communicates directly with the effector. The more complex this nervous chain, the more intricately may nerve impulses be relayed through different paths in the body, but in all cases the net result is the same. All adjustments are due to the reception of stimuli by sense organs, both internal and external, followed by appropriate reaction of other parts of the body.

The most common type of nervous system among invertebrates consists of a small brain lying above the alimentary tract near the anterior end of the body. See Fig. 5. It is connected by cords passing down around the sides of the gut with a ventral nerve cord. In the annelid worms this cord is a chain of ganglia, one lying in each segment. Conclusive evidence from embryology and minute anatomy shows that the primitive cord was a paired structure and that each segment contained a pair of ganglia connected by transverse nerves as well. In most existent species the adult shows no visible evidence of the paired condition.

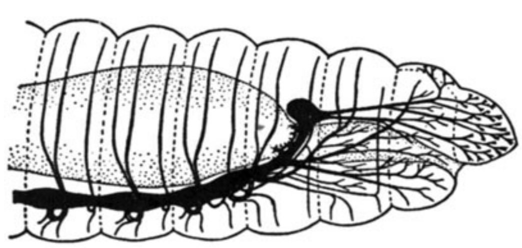


Fig. 5. Nervous system of earthworm. This is a lateral view of the arrangement of the larger nerve trunk in the left side of the anterior segments of the earthworm.

In all of the more complex nervous systems, a major center, the brain, is present as a part of the central nervous system, and supplementary centers called ganglia, each containing a small number of nerve cells, are scattered through the outlying parts of the system. All of these parts are connected by nerve cords or nerves and are joined in the same way with other structures of the body.

In the arthropods a concentration of ganglia in some species has resulted in one large ganglionic mass near the anterior end of the ventral cord.

Many mollusks have a number of ganglia of about equal size as nerve centers, although the cephalopods have a concentrated brain equal to any other invertebrates.

In the vertebrates the entire central nervous system lies above the alimentary tract. The brain is large and complex in most classes. From it a spinal cord runs back along the axis of the body. Both brain and spinal cord bear nerves which extend throughout the animal. See Fig. 6. Below the spinal cord two chains of ganglia connected by slender nerves with each other and with the spinal nerves constitute the sympathetic chains of the autonomic nervous system. Nerves of this system supply the viscera and are widely distributed elsewhere in the body. In cooperation with the parasympathetic fibers of the cranial and sacral nerves, they control the functions which cannot be deliberately regulated by individual desire, such as the movements of the alimentary tract.

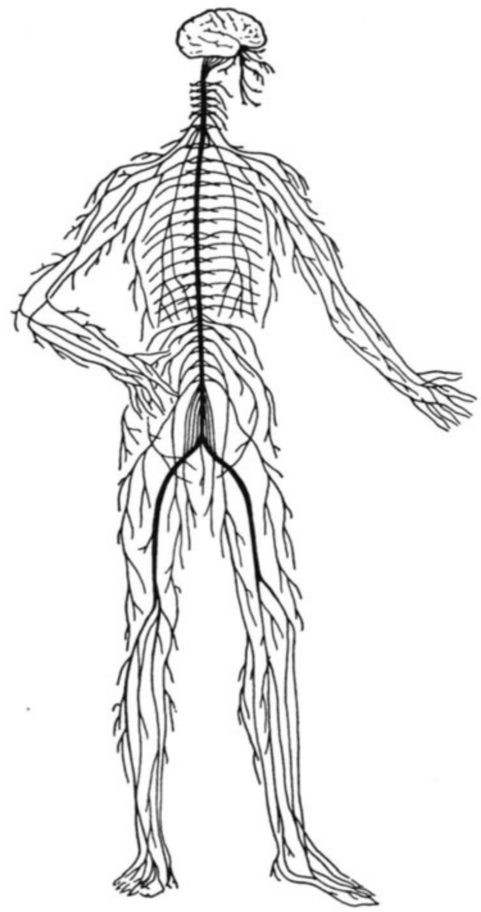


Fig. 6. Human nervous system.

Early Nerve Development in the Brain. Although poorly understood, the brain appears to undergo at least two developmental (constructional) phases. Sometimes referred to by scientists as "temporary scaffolding" (comparable to construction aids for mechanical structures, such as a building), researchers suggest that these *subplate* neurons serve as a temporary linkage for the forthcoming permanent adult brain structure. The early brain is quite different from the mature brain. The analogy has been suggested that in some instances a group of short-lived cells may serve as intermediate targets and "tour guides" for incoming axons. (See Holloway reference listed.)

Synapses and Neuromuscular Junctions. The area of functional contact between two nerve cells (*synapse*) and between an axon terminal and a muscle fiber (*neuromuscular junction*) are regions of unique anatomical, physiological, and biochemical specializations that are concerned with information transfer. Since there is a wide variety of types of synapses and neuromuscular junctions, to date investigations have been relatively limited to mammalian, amphibian, and invertebrate species of select categories.

Synapses and neuromuscular junctions have been classified on a variety of morphological, functional, and biochemical bases. Anatomically, synapses have been classified as being either central (brain and spinal cord); or peripheral (autonomic ganglia, peripheral sensory ganglia and effector junctions) in position and/or axo-dendritic, axo-somatic or axo-axonic in contact relationships. Physiologically, synapses have been classified as either excitatory or inhibitory in action and/or as chemically or electrically mediated. Biochemically, synapses have been classified as either releasing acetylcholine, epinephrine, norepinephrine, or other substances that are currently being researched. Neuromuscular junctions are classified anatomically as within the peripheral nervous system; functionally as excitatory, inhibitory or sensory; and biochemically as releasing acetylcholine.

When synapses and neuromuscular junctions are observed with the electron microscope, their fine structures are seen to be basically similar. The similarities include increased electron density of the presynaptic and postsynaptic membranes, the presence presynaptically of large numbers of mitochondria and synaptic vesicles, and an intersynaptic cleft space of 200–300Å. Synaptic vesicles are 300–600Å in diameter in synapses where acetylcholine is the mediator; and 600–1000Å in diameter in synapses where epinephrine or norepinephrine are the mediators. The accumulation of mitochondria and synaptic vesicles on the presynaptic side occurs as a result of the distal migration of materials from the cell body into the axon. Differences in the morphology of central nervous system synapses have been reported and various functional attributes have been alluded to. These include differences in presynaptic and post-synaptic interspace distances, postsynaptic membrane electron densities, and amounts and electron density of interspace substance.

Synapses transmit either electrically or chemically, and they may be either excitatory or inhibitory in action. Most synapses and neuromuscular junctions are chemically mediated and excitatory in action. Types of chemically mediated synapses include those from the mammalian and amphibian central nervous system, sympathetic and parasympathetic ganglia, and sympathetic innervation of the adrenal chromaffin cells. As previously mentioned, the three most widely known transmitter substances are acetylcholine, epinephrine, and norepinephrine.

Operation of the central nervous system depends upon two substances: the *gray matter* (Nerve cells), and the *white matter* (nerve fibers given off-by the cells). The function of the gray matter is the generating and dispatching of nerve impulses. The function of the white matter is the conduction of these impulses to and from the cells in the gray matter. Other cells in the nervous system have no nervous function, but instead are concerned with the support and nourishment of nerve cells.

Nerve tissue proper consists of cells giving off threadlike processes (*axons*) some of which are extremely long. The axons connect with other cells in the brain or spinal cord. These cells constantly generate, receive, or store up energy. Unlike most body tissues, nerve cells are not replaced once they are destroyed. If the cells are destroyed, their axons degenerate. Some cells concerned with generating or receiving similar impulses may be collected into definite groups called nuclei. The axons from these cells unite and form bundles of nerves which then transmit the impulses: The nerve cells are not collected into nuclei in the cerebral hemispheres, but form a uniform layer (*cortex*) of gray matter over the surfaces.

A series of highly specialized organs, called *receptors*, detect changes in and about the body. They rapidly transmit this information to definite stations within the nervous system. This is called *sensory* activity. Receptors, called *exteroceptors*, gather information from a distance, such as seeing, hearing, and smelling. *Interoceptors* detect things in contact with the body, as pain, temperature, and touch. *Proprioceptors* pick up information from within the body, giving a sense of bodily position. Fibers from the receptor organs pass into the spinal cord as a part of the nerve. Within the cord, they unite to form ascending tracts which connect with other spinal cells or enter the brain. Fibers from some special receptors, such as the eye, form nerves which enter the brain directly.

It is postulated that when a sensory impression reaches the brain, it stimulates a nerve cell which, in turn, stimulates another cell. A third cell is then stimulated and so on, until a circle has been completed, and the last cell restimulates the first one. The circuit continues to *reverberate*, thus retaining the impression which set it off. It is further postulated that these reverberating circuits hold the impressions so that they can be recalled later, or compared with other impressions. It is believed that a cell may participate in more than one circuit, thus accounting for various associations of sensory and muscular activity.

Synapse Modification. In a 1987 paper, Bear, Cooper, and Ebner (Brown University) reported on a physiological basis for a theory of synapse modification. The researchers point out that the functional organization of the cerebral cortex is modified dramatically by sensory experience during early postnatal life. The basis for these modifications appears to be a type of synaptic plasticity that may also contribute to some forms of adult learning. The question of how synapses modify according to experience has been approached in the reported study by determining theoretically what is required of a modification mechanism to account for the available experimental data in the developing visual cortex. The resulting theory states precisely how certain variables may influence synaptic modifications. This insight has led to the development of a biologically plausible molecular model for synapse modification in the cerebral cortex.

Neurons

No two nerve cells (neurons) are exactly the same, but the majority of neurons as presently understood have three main structural features: (1) The *cell body*; (2) the *dendrites*; and (3) the *axon*. See Fig. 7.

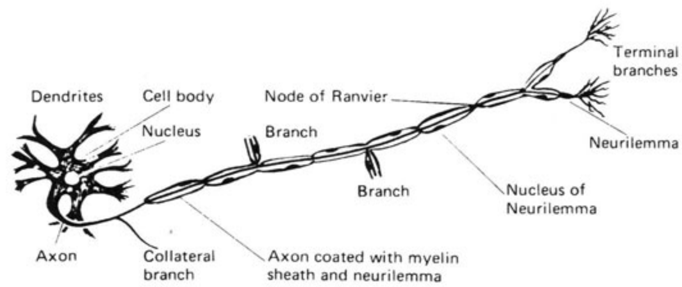


Fig. 7. Simplified presentation of principal elements of a neuron (typical of a motor neuron from ventral column of spinal cord).

Cell bodies may be spherical or pyrimidal in shape. They contain the nucleus of the neuron, that is, the biochemical substances and associated processes required to synthesize enzymes and other substances needed to maintain the life of the cell.

The dendrites (*dendrons*) most often are short and comparatively thick at their points of origin and branch out in a treelike manner. There is a variable number of dendrites with various cell bodies, ranging from a few to many in number. Generally, the dendrites are considered to be the information-receiving channels of the neuron.

The axons are threadlike channels which connect the cell body and dendrites to remote motor and other cells. In some cases, an axon may be as much as half as long as the entire human body. Generally, the axons are considered to be the information-transmitting channels of the neuron. The axons taper very little as they progress outward from the cell body, but frequently they have multiple branches. These branches are called *collaterals*. The axons are filled with a jellylike fluid called *axoplasm*. This fluid was once thought to be simply an inert mechanical support for the excitable membrane that propagates the nerve impulse. Research commenced by Weiss in the late 1940s led to the modern concept of axoplasm, which likens it to an artery for bidirectional flow of molecules moving between the cell body of the neuron and its axon terminals. A number of different systems apparently are operative within the axoplasm. For example, there is a comparatively slow system (movement of about 1 millimeter day) which accounts for the greatest volumetric flow of substances in the axonal system. These are materials required for the growth and regeneration of the axon. There is a much faster system which transports cellular components, such as enzymes, specifically required for the manufacture of transmitters. The different rates of axonal transport are poorly understood, but in essence appear to be related to rates of demand.

The importance of axonal transport to neuron maintenance is revealed by the fact that once the central nervous system is fully developed, there is no replacement of old cells with new ones as occurs, for example, in the liver. Therefore, it is mandatory that the original cells be well maintained for long life.

While attention is usually concentrated on the neurons when describing the brain and central nervous system, it is important to mention in passing the dense network of blood vessels that furnish oxygen and other nutrients to the nerve cells. Connective tissue is also required, particularly at the surface of the brain. There are also *glial cells* which take up nearly all the volume in the nervous system that is not required by the neurons. The detailed functions of the glia are not well understood, but it can be safely generalized that the glia provide both structural and metabolic support for the intricate network of neurons.

Returning to the axons—these are jacketed by Schwann cells, a process which commences during embryonic development. A multiple layer of these cells develops and insulates the axon. This material is called *myelin* and the insulating coating is called the *myelin sheath*. It should be mentioned that some neurological disorders arise from disturbances of this sheath. In most axons, the myelin sheath is interrupted at intervals of about 1 millimeter or more. These interruptions are called the *nodes of Ranvier*. It is believed that these interruptions make it possible for extracellular fluid to directly contact the cell membrane and possibly serve as a source of certain membrane proteins, which are described later. Researchers believe that the myelin sheath functions to conserve the metabolic energy of the neuron. Research has shown that nerve fibers with the myelin sheath conduct nerve impulses at a faster rate than unmyelinated fibers. It will be noted from Fig. 7 that some parts of the axon may be protected only by a substance known as *neurilemma* (where the myelin sheath is absent).

The importance of conserving the metabolic energy requirements of the neuron is demonstrated by the fact that, even though the human brain constitutes only about 2% of total body weight, the brain and central nervous system require at least 20% of the body's total oxygen needs when the body is at rest. This represents a use of oxygen at a rate of about 50 milliliters per minute. The brain metabolic rate is essentially constant even when the body is at rest, and some researchers suggest that during certain phases of sleep, the rate of oxygen required may rise. The total energy equivalent of brain metabolism is about 20 watts. Unlike some other cells of the body which can utilize a variety of fuels, such as sugars, fats, and amino acids, neurons require blood glucose exclusively. Since the brain depends entirely upon oxidative metabolism, consciousness is lost within 10 second and brain tissue degeneration commences upon cessation or severe dimunition of oxygen supply and blood glucose. Diabetic coma from overdosage of insulin, which greatly lowers blood glucose levels, is exemplary of this dependency. More detail on brain metabolism is given a bit later.

Neuronal Information Exchange. Although the detailed functioning of the neurons remains somewhat vague, it is known that a typical neuron may exchange information with a thousand or more other neurons. These connections (synapses) may arise between: (1) The axon of one neuron and the dendrite of another (apparently this occurs in the majority of cases); (2) the axon of one neuron and the axon of another; (3) the dendrite of one neuron and the dendrite of another; and (4) the axon of one neuron and the cell body of another. These four possible connections are probably not fully representative. It is this great connective potential among millions of neurons that leads to connective networks that are unknown even to the most complex of electronic circuits designed by humans and that, from the standpoint of traditional electronic connective networks appear unworkable. The fact, of course, is that they do work.

Neuroscientists are becoming increasingly interested in the role of NMDA (N-methyl-D-aspartate) receptors in the regulation of many normal changes in the brain of an adult or developing animal. Also, there is increased awareness that NMDA may be involved in nerve cell death as encountered in some heart attack and stroke patients. Currently it is believed that NMDA does not exist in the normal brain, but there are NMDA receptors that can be excited by NMDA when used as a tool to probe the activity of this class of receptors. It has been learned, for example, that amino acid neurotransmitters, such as glutamate or aspartate, may stimulate the NMDA receptors.

Charles Stevens (Yale University of Medicine) observes, "Excitotoxicity comes into play if there is too much glutamate and too much calcium. It is not known what the actual mechanism of cell death is, but a current guess is that calcium stimulates protein- and fat-digesting enzymes within the cell that destroy it." Researchers are now seeking a new class of compounds that may block the activity of the receptors and thus prevent nerve cell death. Stuart Lipton (Harvard Medical School) has suggested that, in the brain of a stroke patient, damage to nerve cells increases with the release of more and more excitatory neurotransmitters. The drug MK-801 may be effective in blocking the effects in a patient with escalating levels of excitatory neurotransmitters in the brain. Enrichment of the literature in this area is increasing at a rapid rate.

At a point of connection (synapse), the axon normally enlarges and takes the form of a *terminal button*. This button incorporates tiny chambers (*synaptic vesicles*) which contain a *chemical transmitter*. For many years, the concept that a neuron released only one transmitter chemical from all its terminals—as proposed originally by Sir Henry Dale—was considered inviolable. Within the past few years, researchers have found that a number of neuropeptides can coexist in the same neurons as norepinephrine or serotonin. The functional significance of such a dual-transmitter system has not been established. The quantity of chemical transmitter released by the terminal button vesicles can aptly be described as vanishingly small. The quantities range from a few hundred to a few thousand molecules. As previously mentioned, transmitter chemicals may be excitatory or inhibitory.

Neuron Membranes. Like other cells of the body, neurons have membranes. In the neuron, the membrane is made up of two layers of lipid molecules. The membrane is about 5 nanometers (~0.0003 millimeter) thick. The hydrophilic ends of these molecules are oriented toward the water on the inside and outside of the cell. The hydrophobic ends are oriented away from the water and thus these form the interior of the membrane. Lipid materials tend to be relatively uniform for all cells. But membranes do differ, and these differences arise from the specific proteins that are associated with a given membrane. Membrane differences comprise just one more way in which neurons differ from each other, adding to the overall complexity of classifying and understanding them.

When the proteins are embedded in the bilayer of lipids, they are called *intrinsic proteins*. Although essentially bound, in some cells these intrinsic proteins can move about a bit by diffusion, while in other instances they are physically fixed to the membrane at specific sites. Other proteins, equally important, attach themselves to the membrane surfaces, but are not an actual part of the membrane structure.

These membrane proteins fall into at least five categories, based on their function: (1) *Pumping proteins* maintain the needed concentrations of particular molecules within the cell and thus set up the conditions necessary for a nerve impulse. The pumping proteins regulate concentrations by moving ions and molecules against concentration gradients and hence require considerable metabolic energy. (2) *Channel proteins* provide pathways (tiny orifices) through the lipid bilayer membrane and thus regulate the diffusion of specific ions through the membrane because the ions are of discrete size, some larger than others. Without these channels, the membrane would be essentially impervious. (3) *Receptor proteins* provide binding sites to enable the cell membrane to recognize and attach various specific kinds of molecules. The receptor proteins are highly specific, much as a lock is specific to one key. (4) *Enzyme proteins*, located in or on the membrane, participate in chemical reactions at the membrane surface. For example, one of these enzymes, known as adenylate cyclase, regulates the intracellular substance cyclic adenosine monophosphate (cylic AMP). This substance serves both as a regulator and catalyst. (5) *Structural proteins* contribute to the physical support of the system.

Membrane proteins are keys to the function of the neuron because they provide the means for creating and propagating the nerve impulse. The concentrations of sodium and potassium ions on either side of the cell membrane differ markedly. In a resting state, researchers have estimated that this difference in ionic concentrations causes the interior of the axon to be about 70 millivolts negative with respect to the exterior. In terms of the sensitivity of modern instruments, this is a very significant voltage differential. As a nerve impulse propagates along the axon, the permeability of the axon membrane to both sodium and potassium ions is altered. The impulse is usually triggered in the cell body as the result of reaction to dendrite synapses, and the voltage across the membrane just ahead of the traveling impulse is lowered. This adjustment of sodium ion channeling permits sodium ions to flood into the axon. During this process, the internal potential of the membrane is changed from negative to positive. But the sodium channels remain open for but an instant, after which the potassium channels are opened to permit the outward flow of potassium ions, thus returning the original differential of −70 volts to the interior of the membrane. The sharpness of these actions appears as a spike on an oscilloscope and the phenomenon is generally called *action potential*. See also **Action Current**; and **Action Potential.** In actuality, since the axon is divided into numerous sections (Fig. 7), a wave of voltage propagates along the axon until the terminal button is reached. The action is analogous to a flame traveling along a fuse to an explosive. An explanation of this action in traditional terms is given in Fig. 8.

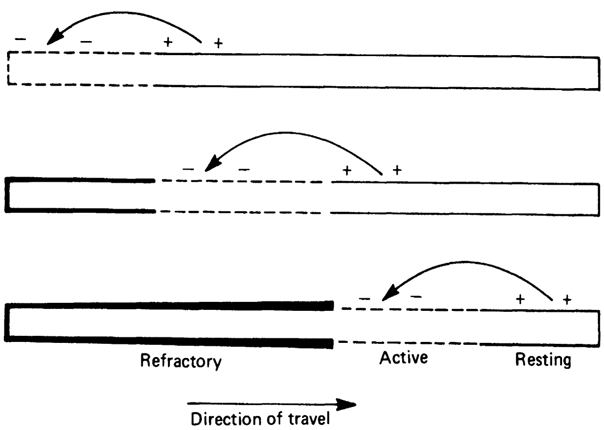


Fig. 8. Simplified representation of nerve impulse conduction.

As previously mentioned, the channel proteins accommodate the sizes of specific ions. The sodium ion is about 30% smaller than the potassium ion. Some channel proteins are relatively undiscriminating, permitting only a slightly reduced number of sodium ions versus potassium ions to pass through the channel. Other channel proteins are highly selective, allowing, for example, fewer than 10 sodium ions to pass for every 100 potassium ions. The diameter of the former type of channel is about 0.8 nanometer and the pore is filled with water. The opening of the second type of channel is considerably smaller and, of course, contains less water. These variations again exemplify the rather extreme degree to which neurons can be tailored to need and thus serve not simply as on-off switches, but provide shading, averaging, and tempering of input and output. The voltage-gated channels just described are traditionally named sodium or potassium channels in accordance with which ion is most readily passable. In other parts of the nervous system, calcium ions also play an important role.

The measured potential (70 millivolts) is less than the equilibrium value calculated by the Nernst equation for the potential that would be developed by a simple concentration battery formed by the unequal distribution of sodium and potassium ions. The reduced potential is probably the result of special back-pumping of the sodium (extruded) ions and the potassium (reabsorbed) ions against their concentration and electrochemical gradients. This pumping action proceeds simultaneously with the diffusion and thus distorts the picture of diffusion as would be determined by membrane permeability alone.

The resulting modified potential is the source of current that eddys into the point of dropped (or even reversed) potential resulting at the site of stimulation, permitting temporary free diffusion of ions and a resulting depolarization. The transient free diffusion of ions across the cell membrane carries the current into the cell. The eddy currents, in turn, act as stimuli to the surrounding membrane, which is then depolarized and draws current from more regions adjacent to the new stimulus sites. These new currents, in turn, stimulate more distant regions, and so on. Meanwhile, the original sites of stimulation and depolarization are repolarizing or recovering.

These events are shown in simplified format in Fig. 8, showing the change of resting, polarized to active, depolarized region, and the influence of the subsequent eddy current in creating a progression of the active region from left to right, tailed by a region of restored membrane. The latter is temporarily more stable than normal and, therefore, resistant to restimulation, that is, it is refractory for a short period. The eddy of current in this way spreads into a self-propagated wave, which is the conducted impulse whose minute electrical accompaniment can, after suitable amplification, be recorded as the action potential.

The channelling characteristics of the membranes of axons vary considerably, depending upon specific function. There are two principal forms—in one type, a nerve impulse opens and closes the channels in response to voltage differences across the cell membrane. Channels of this type are said to be *voltage gated*. In the other type, there is little if any response to voltage changes, but instead they respond to a particular chemical substance, as when a transmitter substance binds to a receptor region on the channel protein. These are called *chemically gated*. It has become traditional to name the chemically gated channels in accordance with the transmitter chemical involved, such as a GABA-activated channel.

In a 1986 paper, J. Dumont and R. Robertson (McGill University) suggested that neuronal circuits be considered from an evolutionary perspective. They point out that to understand neural circuits fully, it is necessary to know not only how they work, but also why they work that way. Answers to the latter questions have been almost teleological in their assumption of optimal design. However, close examination of certain systems has revealed features that apparently lack adaptive value. Their existence can be understood only if the evolution of these circuits is considered and, in particular, how nonadaptive determinants have guided their evolution. In making their proposal, the researchers describe the neuronal circuitry of the locust and the crayfish. Whether a study of arthropods can be translated to human nervous systems is uncertain.

In a scholarly paper (1992), Bert Sakmann (Max Planck Institut für Medizinische Forschung, Heidelberg, Germany) discusses the fundamentals of the plasma membrane of a cell that separates its interior from the extracellular environment and from other cells and acts both as a diffusional barrier and as an electrical insulator, allowing differentiation of cells with specialized functions. He generally reviews how the nervous system connects cells in a very specific way, including the signal transmission between individual cells that take place at contacts (synapses) and that are anatomically and functionally highly specialized. Mention of relatively recent experimental methods, such as the patch-clamp technique, are described in the article. Sakmann was awarded the 1991 Nobel prize in physiology or medicine. Copies of the lecture are obtainable.

Patch-Clamp Technique. This method for separating ion channels in cell membranes was introduced by Sakmann and Erwin Neher in the mid-1970s. Since that time, several hundred research laboratories have adopted the technique. Although quite simple conceptually, several years were required to refine the technique.

Basically, the technique requires a specially contoured, extremely thin glass pipette, which is sealed tightly against a cell membrane. Thus, a small patch of the membrane and the ion channel that it contains can be isolated. Once isolated, the channels can be manipulated chemically or electrically. In some cases, it may be possible for an investigator to construct a window in a living cell for the purpose of altering the cytoplasmic constituents. This provides, in essence, a "hands on" approach to cytology at the microscopic level. The objectives of most patch-clamp technology are to probe how ion channels affect membrane voltage and to observe cell processes, including secretion and contraction.

As early as the late 1960s, researchers showed that, when trace amounts of certain antibodies or proteins are inserted into the cellular membrane, they become electrically conductive. Neher and Sakmann [see paper listed at end of article] observe, "The discrete changes in the current passing through the membranes suggest that the proteins created porelike channels that opened and closed individually. Charged ions then can traverse the membrane through the open channels."

The neuromuscular junction of skeletal muscle has been the object of numerous experiments. Here the presynaptic neurons release acetylcholine in discrete multimolecular packets (quanta). Molecules of acetylcholine transiently bind to acetylcholine specialized proteins in the post-synaptic membrane, thus creating a current flow across the end-plate membrane. End-plate current is the sum (aggregate) of elementary currents through several hundred thousand channels. As pointed out by Neher and Sakmann, "To increase these elementary currents individually, one can press the top of a patch pipette onto a muscular fiber at the end-plate membrane. This region of muscle surface contains the acetylcholine receptor channel. When a low concentration of acetylcholine is in the pipette solution, the current recorded with the pipette switches between two levels. At one level, effectively no current flows because all the ion channels in the membrane patch are shut. When one channel molecule flips to the open state, because of a voltage applied to the membrane, a current of about 2.5 picoamperes abruptly flows through it. After a variable period, the molecule flips back to the shut state, and the current switches off."

The patch-clamp technique, in addition to illuminating molecular details of channel function, also is useful in the study of signaling mechanisms at the molecular level by way of a technique referred to as *voltage clamp analysis,* the development of which is attributed to K. S. Cole (Marine Biological Laboratories, Woods Hole, Massachusetts) as early as 1949.

In summarizing the advantages of their patch-clamp technique, Neher and Sakmann observed in 1992, "The extreme sensitivity of the patch-clamp technique has revealed molecular details of how channels function. The technique has enabled us to study the miniscule cells of mammalian tissue and to trace their signaling pathways by controlling the intracellular environment. Although it is powerful, the technique is also very simple. As more researchers adopt it, we hope and expect that it will continue to help unlock cellular secrets."

It should be mentioned that the thin glass pipette is approximately $\frac{1}{25,000}$ the diameter of a human hair.

Chemical Messengers in the Nervous System. These messengers can be placed into three overlapping categories according to their mode of action—as suggested by R. H. Scheller, et al. (Stanford University): (1) *neurotransmitters*, which act quickly and over short distances to alter transiently the excitability of cells; (2) *neuromodulators*, which alter the response of a cell to a distinct chemical input; and (3) *neurohormones*, which act at a distance from their site of release and give rise to neuronal effects that tend to be slow in onset. In studying chemical messengers, Scheller and colleagues have observed the central nervous system of the gastropod mollusk *Aplysia*. The marine organism can reach 30 to 50 cm in length and often weighs up to 1 kilogram. The central nervous system contains some 20,000 neurons clustered into 5 major ganglia, which are interconnected by an intricate network of connective nerve tracks. The research team has applied recombinant DNA techniques to isolate the genes that encode the precursors of peptides expressed in identified neurons of known function. The organization and developmental expression of these genes have been examined in detail. Several of the genes encode precursors of multiple biologically active peptides that are expressed in cells which also contain classical transmitters. These studies, as well as immunohistochemical studies and the use of intracellular recording and voltage clamp techniques may be the first steps toward revealing the mechanisms by which neuropeptides govern simple behaviors.

Some of the better known chemical transmitters, found to be monoamines or amino acids, are illustrated in Fig. 9. The most common of inhibitory transmitters found in the brain is gamma-aminobutyric acid (GABA). This substance is produced almost exclusively in the brain and spinal cord. Some researchers believe that about one-third of the synapses in the brain involve GABA as a transmitter. It has been reasonably well established that a deficiency in brain GABA is the cause of Huntington's chorea. See **Chorea (Huntington's).** To date, it has not been possible to replace deficient GABA in the brain because no GABA analogues thus far developed can penetrate the blood-brain barrier. See **Blood-Brain Barrier.** Recent research has indicated that the effectiveness of certain anti-anxiety drugs, such as diazepam (Valium®) and other drugs of the benzodiazepine class, may derive their effectiveness as psychoactive drugs because of their ability to increase the effectiveness of GABA at its receptor sites.

Amino acids, such as glycine, GABA, and L-glutamate, are the most prevalent signaling substances for rapid communication. Channels that bind these transmitters open and close randomly. Thus, these receptor channels operate in a manner similar to the acetylcholine channels at the end plate. Neher and Sakmann observe, "Transmitter-gated channels in the central nervous system often show an additional complexity, in that some channels may be only partially open or closed and that different sub-types of channels may occur in various brain regions."

A better understanding of the manner in which chemical transmitters are manufactured and utilized is developing, with the greatest progress taking place during the past decade. Several steps are involved in the production of these substances. One or several enzyme catalyzed steps

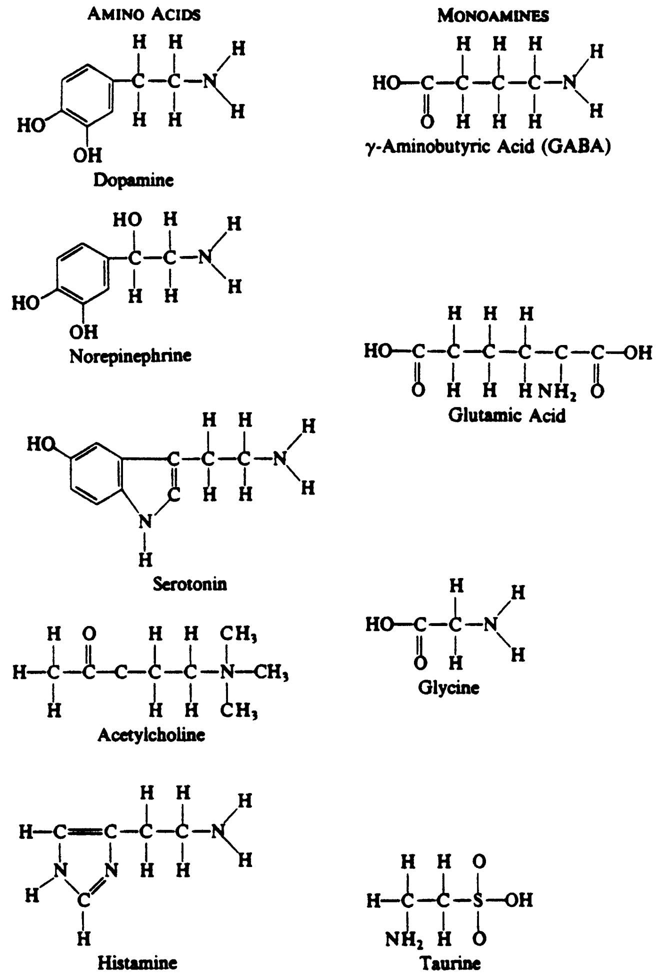


Fig. 9. Neurological system transmitter chemicals. Typically, these substances are small molecules that contain a positively charged nitrogen atom. Usually each substance characteristically inhibits or excites neurons, but in some instances the transmitter substances may play one role in one part of the brain and the other role in another part of the brain. At one time it was believed that all terminals of a given axon will transmit only one substance, but there is growing evidence that this may not always be true.

may be required to transform precursor molecules (which are stored in readiness for conversion) to the final transmitter chemicals. In the instance of acetylcholine, only one enzyme-catalyzed step is required. In the cases of some others, such as norepinephrine, three such steps are required, commencing the synthesis with tyrosine, which is taken up by the nerve terminal from the bloodstream. Tyrosine is converted to L-DOPA. This intermediate is then converted to dopamine and, in another step, dopamine is converted to norepinephrine. Once manufactured, vanishingly small quantities are temporarily stored for a short period (until needed) in the synaptic vesicles. The tiny quantities involve only between 10,000 and 100,000 molecules of transmitter substance. These molecules in the vesicles are protected from enzymes in the terminal that would degrade them. (The mechanism of protection is not fully understood.) A nerve impulse triggers the expulsion of the transmitter substance from the synaptic vesicles into the synaptic space; this transfer in turn activates other neurons. Few details are known concerning the transfer mechanism, but it has been established that an increase in the permeability (to calcium ions) of the nerve terminal serves to engage the release mechanism. Once released, the transmitter molecules seek out specific receptor sites and thus propagate the effects of a single nerve impulse.

Although a minor degree of interaction can take place between the electrical fields of adjacent fibers, the neural impulse conduction system is essentially an adequately well-insulated system, suited to discrete communication. The development of widespread patterns of activity reflects the functioning of interconnections at the synaptic level. It should not be surprising, since transmission is effected by specific chemicals, to find that under special circumstances of great stress, a major depot of such chemicals (adrenaline and noradrenaline) located in the medulla of the adrenal gland can pour these synaptic inhibitory chemicals into the bloodstream. Distributed in this way to all synapses, the adrenal medullary secretion can supply a cutoff influence limiting the massive discharge of impulses that stress is apt to initiate, and thus prevent it from becoming so excessive that it is detrimental. In this way, the neurohumoral transmission mechanism also affords a means of chemical homeostatic regulation.

The vulnerability of impulse transmission to chemical influences may not always serve homeostasis; for here also is where the action of poisons like mescaline and lysergic acid diethylamide takes place. They produce in humans a temporary mental derangement or psychosis. Experiments show that these substances inhibit impulse transmission at cerebral synapses and that tranquilizers prevent this effect. Such experiments in disturbed impulse transmission are developing the basis for understanding of chemically-induced psychosis and the manner of action of tranquilizers. A sufficient parallelism appears to exist between the experimental laboratory findings and data in clinical psychosis to suggest that the latter may, also, sometimes be a disturbance of impulse transmission in the brain and that tranquilizers tend to restore synaptic equilibrium.

The Neuropeptides. During the past decade, the number of chemical-messenger systems identified has increased dramatically. This advance is highlighted by the discovery of a large family of brain chemicals known as *neuropeptides*. As shown by Fig. 10, these molecules are made up of long chains of amino acids, ranging from a few to as many as 39 amino acids. Research indicates that these substances are resident within the neurons. A few of these substances, such as corticotropin (ACTH) and vasopressin, have been known for many years and identified with the hypothalamus and pituitary gland. Probably of greatest interest to researchers are the *enkephalins* and *endorphins*. These chemicals are strikingly similar to the opiate morphine. Identification of these substances followed the finding that specific regions of the brain possess receptor sites that bind opiates with a high affinity. These sites were revealed through the use of radioactively labeled opiate compounds. Further research showed that opiate receptors are located in those regions of the brain and spinal cord which are known to be associated with pain and emotion. Pioneering research efforts in this field included the work of S. H. Snyder and C. B. Pert (Johns Hopkins University School of Medicine), E. J. Simon (New York University School of Medicine), and L. Terenius (University of Uppsala). The enkephalins (Met- and Leu-), as shown in Fig. 10, were first isolated by J. Hughes and H. W. Kosterlitz (University of Aberdeen) in 1975. Each enkephalin contains five amino acids, one of these being methionine in one case, and leucine in the other case. Shortly after the isolation of these compounds, the endorphins, also morphinelike compounds, were isolated from the pituitary gland. Some researchers have suggested that some of the nontraditional methods used for relieving chronic pain, such as acupuncture, direct electrical stimulation of the brain, and possibly hypnosis, may be effective because these procedures may cause enkephalins or endorphins or both to be released to the brain and spinal cord. The drug naloxone (Narcan®) blocks the binding of morphine and experiments have

Tyr-Gly-Gly-Phe-Met
(Met-ENKEPHALIN)

Tyr-Gly-Gly-Phe-Leu
(Leu-ENKEPHALIN)

Arg-Pro-Lys-Pro-Gln-Gln-Phe-Phe-Gly-Leu-Met-NH_2
(SUBSTANCE P)

p-Glu-Leu-Tyr-Glu-Asn-Lys-Pro-Arg-Arg-Pro-Tyr-Ile-Leu
(NEUROTENSIN)

Asp-Arg-Val-Tyr-Ile-His-Pro-Phe-NH_2
(ANGIOTENSIN II)

Tyr-Gly-Gly-Phe-Met-Thr-Ser-Glu-Lys-Ser-Gln-Thr-Pro-Leu-Val-Thr-Leu-Phe-Lys-Asn-Ala-Ile-Val-Lys-Asn-Ala-His-Lys-Lys-Gly-Gln
(Beta-ENDORPHIN)

Ser-Tyr-Ser-Met-Glu-His-Phe-Arg-Tyr-Gly-Lys-Pro-Val-Gly-Lys-Lys-Arg-Arg-Pro-Val-Lys-Val-Tyr-
Pro-Asp-Gly-Ala-Glu-Asp-Glu-Leu-Ala-Glu-Ala-Phe-Pro-Leu-Glu-Phe
(ACTH, CORTICOTROPIN)

Asp-Tyr-Met-Gly-Trp-Met-Asp-Phe-NH_2
(CHOLECYSTOKININ-LIKE PEPTIDE)

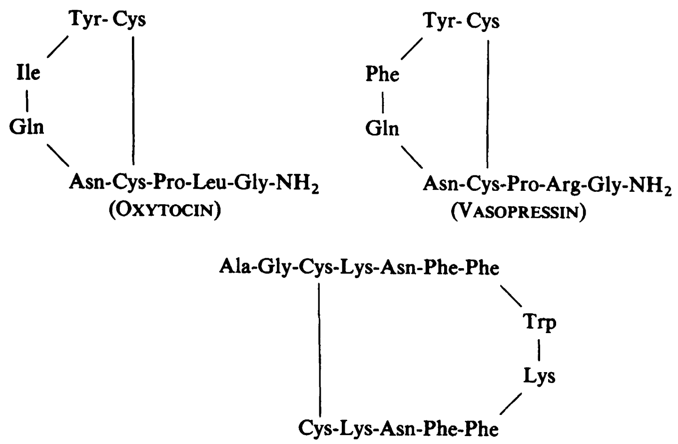
His-Ser-Asp-Ala-Val-Phe-Thr-Asp-Asn-Tyr-Thr-Arg-Leu-Arg-Lys-Gln-Met-Ala-Val-Lys-Lys-Tyr-Leu-Asn-Ser-Ile-Leu-Asn-NH_2
(VASOACTIVE INTESTINAL POLYPEPTIDE, VIP)

p-Glu-His-Pro-NH_2
(THYROTROPIN RELEASING HORMONE, TRH)

p-Glu-His-Trp-Ser-Tyr-Gly-Leu-Arg-Pro-Gly-NH_2
(LUTEINIZING-HORMONE RELEASING HORMONE, LHRH)

Ala-His
(CARNOSINE)

p-Glu-Gln-Arg-Leu-Gly-Asn-Gln-Trp-Ala-Val-Gly-His-Leu-Met-NH_2
(BOMBESIN)



ABBREVIATIONS OF AMINO ACIDS

Ala	Alanine	Leu	Leucine
Arg	Arginine	Lys	Lysine
Asn	Asparagine	Met	Methionine
Asp	Aspartic Acid	Phe	Phenylalanine
Cys	Cysteine	Pro	Proline
Gln	Glutamine	Ser	Serine
Glu	Glutamic Acid	Thr	Threonine
Gly	Glycine	Trp	Tryptophan
His	Histidine	Tyr	Tyrosine
Ile	Isoleucine	Val	Valine

Fig. 10. Now believed to be transmitters, neuropeptides, which are short chains of amino acids found in brain tissue and notably localized in axon terminals, participate in complex mental activity, such as thirst, memory, and sexual behavior.

shown that naloxone also blocks the effects of the aforementioned pain-relieving procedures; hence the tentative hypothesis.

Early research findings indicate that neuropeptides are released from axon terminals through the presence of calcium ions known to be the releasing mechanism in connection with established transmitters. Thus, some investigators believe that the neuropeptides also may be transmitters. This is particularly true of a compound simply identified as *substance P*. See Fig. 10. Some investigators have found that substance P is associated with spinal neurons involved in pain stimuli. P. T. Jessel and L. L. Iversen (Cambridge) had demonstrated that opiate drugs are able to suppress the release of substance P from sensory fibers.

Using experimental animals, researchers have shown that injection of nanogram amounts of a neuropeptide directly into brain tissue (to avoid the blood-brain barrier) causes marked behavioral changes in the subject, such as prolonged drinking behavior (at a time when the animal should not be thirsty), enhancement of sexual behavior, and improvement of memory. Some scientists have suggested, for example, that vasopressin may have beneficial effects on human patients suffering from memory loss.

One of the surprising findings of recent brain and central nervous system research is that chemical substances previously considered to be exclusive to the province of the brain have been found in organs outside the nervous system (examples: somatostatin, neurotensin, and enkephalins found in the gut); and, conversely, that other substances active in other organs, but not previously associated with the central nervous system, have been found in the latter (examples: gastrin, vasoactive intestinal polypeptide (VIP), and cholecystokinin, traditionally associated with the gastrointestinal tract, but now found in the central nervous system).

In an interesting paper (see reference listed), D. I. Gottlieb (Washington University School of Medicine) describes how nerve cells not only excite their neighbors, but also inhibit them. Such inhibitory activity, often mediated by GABA, helps to shape the neural networks that underlie all behavior.

In the early 1980s, some pharmacologists suggested that nitric oxide, a gas, may be an important regulator of body activities. Recent research at the Johns Hopkins University has noted a specific connection between nitric oxide and the brain. In low concentrations, the chemical can act as a neurotransmitter for carrying nerve impulses from one cell to another. In high concentrations, the substance can be toxic to neurons. A factor that impedes research in this area is that nitric oxide has a very short life span; it lasts for only a few seconds after its manufacture. Researchers at Johns Hopkins now are attempting to clone the gene that encodes nitric acid synthase, the enzyme that makes nitric oxide in nerve cells and also in the endothelial cells that line blood vessels.

Astrocytes. The importance of excitable neurons to brain processes is fundamental. However, H. K. Kimelberg and M. D. Norenberg (University of Miami School of Medicine) reported (1989) that approximately half of the brain does not contain excitable cells, but rather is made up of non-excitable cells. The latter are placed in two groups: (1) *neuroglia* (a word first used by a German pathologist in 1846) and sometimes referred to as "nerve glue," and (2) *astroglia,* commonly referred to as *astrocytes*. They are named for their starlike shape. Scientists in the past generally have regarded these cells as passive support elements for the neurons. Improved methods for growing and identifying astrocytes have revealed for the first time the role of these cells in brain physiology, brain development, and nervous system pathology.

Investigators now admit that a better understanding of astrocytes is needed to establish a full picture of the workings of the brain and no longer can be accepted simply as passive in nature.

Once regarded as a cementlike connective tissue, Golgi in the mid-1870s identified *neuroglia* by using staining techniques. A classifying system developed by Santiago Ramón y Cajal (Spanish neuroanatomist) during the same time frame referred to the majority of neuroglia as *macroglia,* a concept that has persisted over the years. Macroglia are made up of astrocytes and *oligodendrocytes*. The latter are known to form myelin, which protects axons and neurons in the white matter of the brain.

Glial fibrillary acidic protein (GFAP) was discovered in the early 1970s as a component of the cytoplasm of astrocytes and became a convenient marker for identifying the cells. GFAP was isolated by L. E. Eng and A. Bignami (Stanford University) in the 1970s. The ability to work with these cells in samples of tissue and in culture led to a further understanding of the role of astrocytes. Research has implicated astrocytes in neuron damage that is manifested by psychiatric disorders ranging from stupor to coma, an example of which is known as *hepatic encephalopathy*. The mechanisms currently involved are poorly understood. Certain psychoactive drugs adversely interact with astrocyte receptors in some still unexplained way. Astrocytes are believed to participate in Huntington's disease, Parkinson's disease, multiple sclerosis, and Raye's syndrome. Researchers Kimelberg and Noremberg observe, "The time has come to consider astrocytes equal partners with neurons in both the normal and abnormal brain."

Protection of Central Nervous System. The total system is well protected by the rigid *skull* and the flexible backbone (*vertebral* or *spinal column*). The skull consists of a dome of thin, porous, but strong bones. The bones of the forehead contain the sinuses of the nose. The floor of the skull is composed of somewhat thinner bone and contains more sinuses that connect with the nasal passages. The deeper structures of the ear are embedded in the bones forming the base of the skull. The floor of the skull consists of three irregular depressions (*fossae*) that form three descending levels. The fossa in the back of the skull is the largest and deepest. The cranial nerves arise from different parts of the brain and emerge through various bony canals and openings of the skull.

Inside this hard shell, three separate tissues provide additional covering for the brain and spinal cord. The outermost, *dura mater*, consists of layers of dense, fibrous material. The outer layer of the dura adheres tightly to the bones of the skull. The inner covers the brain and spinal cord, forming a tough, saclike structure. Within the skull itself, the dura is folded into partitions which separate and support the various parts of the brain. One such fold is called the *falx cerebri*, which divides the cerebral hemispheres into right and left halves. Another fold, the *tentorium*, separates the back fossa from the vault of the skull. It provides a horizontal support for the back part of the brain and separates that part from the cerebellum. Between the layers of the dura are large blood vessels called *venous sinuses* which collect blood from the brain and return it to the heart.

The middle of the three covering tissues is a delicate membrane called the *arachnoid* (cobweb-like), a layer that encloses the brain and the spinal cord in a loose-fitting sack. The space below this layer is the *subarachnoid space*.

The third covering is a thin, delicate sheet that follows closely all the irregular surfaces and fissures of the brain and spinal cord. This is called the *pia mater* and is next to the nerve tissue proper.

Between the arachnoid and the pia mater circulates the cerebrospinal fluid, which acts as a shock absorber for the central nervous system. Also, it is postulated that this fluid helps in nourishing the nerve tissue itself. This clear, watery-appearing fluid, formed within certain cavities (*ventricles*) of the brain, flows out from small openings into the brain and spinal cord before it is absorbed back into the bloodstream. When the flow or absorption of the fluid is impaired, it accumulates in large quantities. The head enlarges and "water-head" (*hydrocephalus*) results. The spinal fluid is affected by various disorders and is easily drawn off to be examined for diagnostic purposes.

The findings of D. Cunningham (University of California, Irvine) pertaining to the role of clotting factor *thrombin* in helping to maintain nerve connections in the brain were reported in 1992. Thrombin is best known, of course, for its role in blood coagulation. Proteins of the body's blood clotting system assist in the regulation of normal brain development and in protecting the brain against stroke damage, trauma, and other injuries. Further, research indicates that an imbalance of these proteins may contribute to nerve-cell damage, as encountered in Alzheimer's and other neuro-degenerative diseases. After years of earlier research, including that of D. Monard (Friedrich Miescher Institute, Basel, Switzerland) in the early 1970s, Cunningham and associates identified newly found protease nexins (PN-1 and 2) as playing an important role in making and maintaining connections between nerve cells. Investigators have found that the action is not only manifested in culture, but also occurs in intact brain tissue. Apparently the actions are limited by the balance between the protein factors. An imbalance can result in protecting nerve connections or, on the other hand, lead to loss of neuronal connections and nerve cell damage, especially if thrombin predominates. It has been suggested by some investigators that tests for the aforementioned substances may become a basis for diagnosing Alzheimer's.

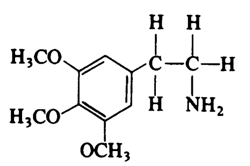
Trophic Substances. Among the most recent discoveries in brain research concerns so-called *trophic substances*, which are believed to be secreted from nerve terminals. One of these substances is nerve-growth factor (NGF). It has been established that this protein is required for the differentiation and survival of peripheral sensory and sympathetic neurons. Considerable research effort is going forward to better understand the function of trophic substances and to isolate and identify them. Substances that may be generated in the pituitary gland and transported to the brain are mentioned in the entry on **Pituitary Gland.**

Another important area of research that benefits from recent findings such as those just described is directed toward a better understanding of how psychoactive drugs interact with the brain and central nervous system. See Fig. 11. An improved understanding will assist in the treatment of various brain disorders and of countering the effects of hallucinogenic drugs that are willfully, but tragically taken by a small number of very foolish people for "kicks." Research has progressively indicated over the years that tampering with the extremely complex brain and central nervous system is indeed attended by many risks.

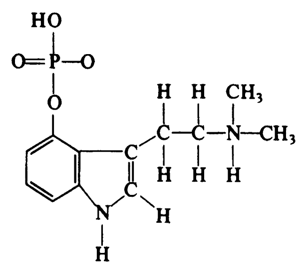
Ethnic Disparities in Drug Therapy. As early as 1974, Keh-Ming Lin (University of Washington) noted that, in treating Asian patients with schizophrenic psychoses in Taiwan, they required much smaller dosages of drugs (haloperidol) than proved to be the case of Caucasian patients. To relieve symptoms among the latter ethnic group required about ten times the dosage (2 mg), as was the case in Taiwan. With further investigation, it was found that Asians often require far lower doses of psychotropic drugs than Caucasians. Prior research with the treatment of such diseases as sickle cell anemia and Tay Sachs disease had indicated differences in response related to the ethnic origin of patients. As of 1991, there are several centers that are specializing in this area, one being that headed by Keh-Ming Lin at the Harbor University of Los Angeles, Torrance, California.

Several genetically influenced biological systems may be involved. It is well known, for example, that an enzyme deficiency limits the ability of Orientals to metabolize alcohol. Studies in the late 1970s by British and Indian researchers compared three groups—Caucasians and Sudanese, each in their adopted residency in Britain, and Sudanese in their native villages. Differences in metabolizing certain drugs were noted, but, in such instances, the diet of the patient appeared to be a dominating factor.

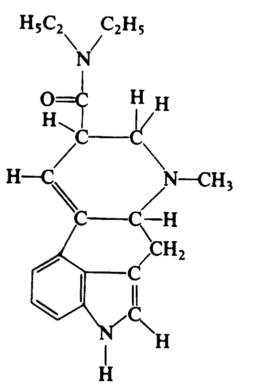
Currently, the role of ethnicity in the use of psychotropic drugs is in a very early stage and one that could lead to controversy. One scientist has observed, "We are not looking at the biology of race differences. We are looking at how different ethnic groups respond to treatment for severe mental illnesses known to have biological correlates. What's important is that a center like this will influence practice directly so people will get more appropriate treatment."



Mescaline



Psilocybin



Lysergic Acid Diethylamide (LSD)

Fig. 11. Representative hallucinogenic drugs which bear structural similarities to some of the monoamine transmitters. It has been hypothesized that these similarities may cause the hallucinogens to mimic natural transmitters at synaptic receptors in the brain. Note presence of benzene-ring structure in these substances, a structure that is present in four out of the five monoamines previously shown. Also note presence of indole ring in psilocybin and lysergic acid diethylamide, a structure that is also present in the monoamines serotonin and histamine.

Brain Metabolism

The metabolism of the brain usually is measured either (1) in living organisms, or (2) in isolated systems, such as brain slices, brain homogenates, or extracts. In isolated systems, the different metabolic pathways can be closely observed, while measurements under living conditions show which of the possible capacities of the organ observed in isolated systems are operative, and at what rate, in the living brain.

The metabolism of living brain usually is measured by analyzing the arterial blood coming to, and venous blood coming from, this organ. Since cerebral circulation is very rapid (under normal conditions, about $\frac{1}{6th}$ of the total circulation goes through the brain), this method can only measure those substances that are used or produced rapidly and in significant quantities by the brain: Glucose is the main metabolite of the brain and the oxidation of this compound accounts for most of the oxygen utilized by the brain. The utilization of oxygen in the brain is surprisingly high, higher than in most other organs, with about one-fifth to one-quarter of the oxygen taken up through the lungs being utilized by the brain, which comprises only 2% of body weight. About 90% of the glucose is oxidized completely to carbon dioxide, the rest mostly to lactic acid, and a small portion to pyruvic acid. The whole brain of a normal young man consumes 46 milliliters of oxygen and 76 milligrams of glucose per minute with a blood flow of 750 milliliters. These figures show that the metabolic rate of the brain is very considerable in the total body economics as far as energy and oxygen utilized, heat produced, and other factors. With physical activity, the oxygen consumption of the whole body changes significantly, but brain metabolism does not. Oxygen uptake and glucose utilization do not change significantly during mental exercise or during sleep, but are altered in extreme circumstances, such as deep anesthesia, which reduces metabolism; or high carbon dioxide content of the air, which increases cerebral blood flow.

The large amount of energy utilized by the brain is somewhat surprising since this organ does not perform work like that done by for example, muscle and kidney. The most plausible explanation is that there is continuous very high activity in the brain. Such activity can be detected by the electroencephalogram. Continuous impulse conduction involves the movement of ions, since conduction by nerve is achieved by temporary change in ion permeability.

Other components metabolized include lipids, nucleic acids, and proteins.

Cerebral metabolic rates change under a number of circumstances. There is a general decrease of oxygen utilization during development. It has been estimated that in the newborn, cerebral respiration can amount to half of that of the whole organism. Metabolism reaches a plateau in adulthood, and further decrease after 50 years of age in humans has been suggested, but not well confirmed. The metabolism of a number of compounds undergoes significant fluctuation under conditions, such as convulsions, hibernation, demyelinating diseases, and other pathological states (multiple sclerosis; myelin). A lot of interest is centered on subtle changes in metabolism which are due to changes in mental activity, such as those shown by comparison of trained versus untrained animals, or animals reared in isolation versus others reared in a complex environment. It is of interest that metabolic changes that were found and which were interpreted as changes in functional metabolism are different in the two types of cells of the brain. In functional metabolism, neurons have preference and glial cells play a supporting role.

The effect of chemical compounds on brain function is well known. Alcohol, antidepressants, hallucinatives all indicate the functional role of metabolism in the brain, particularly of compounds that inhibit cerebral protein metabolism. Compounds that inhibit the latter also inhibit the deposition of short-term memory while most do not appear to affect already deposited long-term memory.

Traditional Views of Memory Function

As of the early 1990s, even though marked progress has been made in recent years, the manner in which the brain processes information (input signals from the various senses), selectively stores that information over short or very long time spans, and retrieves that information upon command remains relatively poorly understood. It appears that not much will be learned from analogies with modern electronic computers, but that the reverse consequence may occur, namely that of using the human brain, once understood, as a model for industrial, business, and scientific computers. Much is being learned currently about the brain by way of improved imaging techniques and computer mapping, subjects discussed later in this article.

Early Research and Hypotheses. Traditionally, the brain has been described in terms of areas, gross structures, and regions, some of which have been identified with memory functions for a number of years. Most of what has been learned to date concerning the physiology of the brain has been acquired from the study of brain injuries and disorders. Memory research is at a very early stage, in which investigators refer to large subsystems—lobes, areas (Broca's and Wernicke's), and other regions with special names—much as a computer scientist refers to the various storage functions and regions of a data processing machine, except that in the latter case, the manner in which the tiny subelements and components of the larger subsystems are well understood. In a way, the most specialized scientists of today, in terms of understanding the brain's memory function, are about in the same position as the layman who has a fair idea of what computers can do, but no depth of knowledge as to how computers actually work. In making this observation, it is not intended that, beyond the general definition of what the animal and electronic memories do, any parallels be drawn between the two fundamentally different kinds of memory systems.

Just as an expert racing driver can quite accurately evaluate the characteristics and capabilities of an automobile, experimental psychologists have collected a large body of data on human learning and memory. To do this, the racing driver need not necessarily know the details of what is under the hood. Thus, we know a lot more about what the memory can accomplish—short-term, long-term, etc.—than we do concerning how these functions are accomplished at the micro or molecular level.

The ability of the human brain to remember in great detail and over long periods of time, thus making language, recognition, music, reading, writing not only possible, but capable of immense refinement, creativity, and imagination—the ingredients of human art, culture, and way of life—is the major distinction of the human being from lower animals.

As early as the 1960s, Broca, a French researcher, showed that damage to a particular region of the cortex will consistently give rise to an aphasia or speech disorder. Now referred to as Broca's area, this is a region on the side of the frontal lobes and is also called the anterior language area. Broca also pointed out over a century ago that damage to this area on the left side of the brain will lead to aphasia, whereas similar damage to the corresponding area on the right side of the brain does not. Over the years, statistics have shown that nearly 100% of all aphasias are caused by brain damage occurring in the left hemisphere of the brain. Frequently, this area is destroyed by a stroke, with at least partial disruption of speech. Somewhat later (1874), Wernicke, a German researcher, discovered another area of the brain associated with aphasia. This area lies between the primary auditory cortex and a structure called the *angular gyrus*. It has been suggested that the angular gyrus mediates between visual and auditory centers of the brain. More recent research also indicates that Wernicke's area and Broca's area are connected by a bundle of nerve fibers known as the *arcuate fasciculus*. These speech centers of the brain obviously must call upon other areas of the brain to bring out all of the stored memory required to recognize, on the one hand, and to formulate on the other hand, the basics of speech and communication.

That the brain functions in what still may be described as mystifying ways is exemplified by a mental disorder known as *prosopagnosia*. Persons with this disorder cannot identify other people's faces at a glance as is true of the normal individual. In fact, persons so afflicted cannot associate other people with photos or with their real faces, but can process and associate voice information held in memory. Other memory retrieval and processing facilities required to read, write, etc. in such individuals are not impaired. Investigators have identified the location of lesions that cause prosopagnosia. Damage is noted on the underside of both occipital lobes, extending forward to the inner surface of the temporal lobes. The present very qualitative state of understanding of the brain's memory and other specialized functions is exemplified by the 1979 statement of a leading American neuroscientist, "The implication (of the aforementioned findings) is that some neural network within this region is specialized for the rapid and reliable recognition of human faces."

Scientists for many years have made qualitative distinctions between short- and long-term memory. There is asymmetry of the right and left halves of the brain. Certain memory processes appear to be associated with structures on the inner surface of the temporal lobe, such as the hippocampus. Bilateral lesions of these areas have been shown to cause a severe and lasting memory disorder characterized by the inability to learn new information. See also **Amnesia**. Patients with lesions of this type appear to have undiminished powers of perception, but they are largely incapable of incorporating new information into their long-term storage. Acute lesions in this region of a single temporal lobe sometimes result in similar, but less persistent memory disorders that reflect the contrasting specializations of the hemispheres. The type of information that can be learned varies according to the side the lesion is on.

Further qualitative descriptions of memory function have been formulated by experimental psychologists. Qualitative descriptions have been prepared concerning visual analyzers and word detectors. For example, some investigators suggest that a person has a visual feature analyzer and collateral sets of word detectors connected to a semantic memory network. It has been proposed that the feature analyzer receives a row of letters and produces a code representing the letter shapes, based upon their lines, curves, and angles, as well as the spatial relations among them. It has been suggested that the code is sent simultaneously to all of the word detectors, each of which takes time to count how many visual features the letter row has in common with a particular word. Qualitative analogues such as these have been very helpful to teachers for improving the practical utilization of the brain's memory and recognition functions, but do not necessarily lead to an accurate concept of the real physiological basis of these brain functions.

In an excellent paper, A. Baddeley (Medical Research Council, Cambridge, UK) describes the conceptual term, *working* memory. The researcher observes, "The concept of a working memory system that temporarily stores information as part of the performance of complex cognitive tasks is proving to be productive. Studies that have utilized the individual difference approach have linked working memory to performance on a range of important tasks, including language comprehension and reasoning. The more analytic approach has shown that the concept forms a useful conceptual tool in understanding a range of neuropsychological deficits, which in turn have thrown light on normal cognitive functioning. Working memory stands at the crossroads between memory attention and perception."

Endel Tulving (University of Toronto) and D. L. Schacter (University of Arizona) define *priming* as a nonconscious form of human memory, which is concerned with perceptual identification of words and objects and which only recently has been recognized as a separate form of memory or memory system. Evidence is converging for the proposition that priming is an expression of a perceptual representation system that operates at a pre-semantic level. It emerges early in development, and access to it lacks the kind of flexibility characteristic of other cognitive memory systems. Conceptual priming, however, seems to be based on the operations of semantic memory.

As pointed out by these researchers, "Memory was traditionally thought to be a unitary faculty of the mind. Recently, however, many researchers have adopted the hypothesis that memory consists of a number of systems and subsystems with different operating characteristics. The problem of what these systems and their properties are, and how they are related to one another, now occupies the center stage in research on memory." The investigators stress the importance of systematic classification of memory systems, both psychological and physiological, as an essential prerequisite for the successful pursuit of the empirical and theorectical understanding of the memory processes and mechanisms. The systems approach, combined with appropriate processing theories, seems to provide the most direct route to the future.

At the 1990 Cold Spring Harbor meeting, debate on the precise cellular basis of memory continued. (See Barinaga reference listed.) For many years, a majority of researchers has contended that long-term memory occurs at the synapses. This is, of course, a generalization that does not satisfy the more curious researchers. Out of further research, the concept that the postsynaptic (receiving) cell is the site of change is emerging. This phenomenon occurs as the result of a process now referred to as *long-term potentiation* (LTP). R. Tsieu (Stanford University) and C. Stevens (Salk Institute) presented details on the LTP concept and the means for researching the process (*quantal analysis*), the details of which are too complex to present here.

At the 1991 Cold Spring Harbor meeting, T. Tully (Brandeis University) and other scientists described their confidence that *Drosophila,* the fruit fly of genetic research, also can become a good model for investigating memory systems, a proposal that received varying responses at the meeting. (See Stipp reference listed.)

A Return to Pavlov. In recent years, interest has returned to Pavlov's conditioned reflex experimentation in dogs. In 1989, D. L. Alkon (National Institute of Neurological and Communicative Disorders and Stroke) reported on research that indicates that a neuron can alter its structure during the process of *associative* learning. Such structural changes previously had been noted only in embryos and immature animals. Laboratory research was done with the alabaster sea slug (*Hermissenda*), which "instinctively" swims toward light. In a programmed experiment, each time the animal observed a light, it was whirled on a turntable, causing the slug to grip the chamber wall tightly by clenching its foot muscle, thus altering the animals' normal response to light. Careful procedures were followed so that any changes observed in the trained group were related specifically to the learning of a new response. Subsequently, the researchers examined a particular neuron (type B photoreceptor, one of five light-detecting neurons in the slug's eye). The cell from a slug with conditioned response was stained and compared with the similar feature of an unconditioned slug. A significant difference was noted. In the trained animal, the axon spread out in a fanlike shape over a much smaller region (50% or smaller) than the axons taken from two control groups. Animals that learned to respond most strongly had the most tightly focused axon terminals. Researcher Alkon observes, "We see the same biochemical and molecular changes in the rabbit hippocampus as we've seen in *Hermissenda*. We'll be looking for the same structural changes as well."

Imaging the Brain

Since the late 1980s, dramatic progress has been made in imaging the brain, both from the objectives of (1) creating still maps and (2) of observing the brain at work, in connection with research and diagnostic procedures. Principal imaging techniques used have included (1) x-ray computed tomography (CT), (2) positron-emission tomography (PET), (3) magnetic resonance imaging (MRI) and (4) single-photon-emission computed tomography (SPECT). Although not an imaging technique per se, much nervous system information also is obtained by way of magnetoencephalography (MEG).

Much research in the past has been obtained by using invasive research techniques on animals, notably the macaque monkey. Imaging techniques now are used extensively in animal research. One investigator has developed a chart that reveals 32 areas of the macaque's cerebral cortex associated with vectors of vision, including 300 circuits that connect these areas. The detail is much greater than currently is obtainable in examining the human brain, which is ten times larger than the macaque brain.

Prior to sophisticated imaging processes, human brain research essentially was confined to gleaning information from limited and rather crude methods, such as studying brain lesions and making electrical recordings of neural activity in patients while undergoing brain surgery.

X-Ray CT. These images result from the passage of x-rays through the brain or spine in a number of different directions. Intervening tissues attenuate radiation to different degrees, thus producing two-dimensional multiplanar representations of the structures under examination. Detectors in opposition to the series of x-ray sources measure the amount of transmitted radiation. Computer algorithms relate the density of the intracranial or spinal contents to the quantity of attenuation. These are are known as tomographic images. By use of iodinated contrast agents given intravenously or intrathecally, tissue attenuation can be increased. The earliest of the imaging techniques, x-ray CT has limitations. Meningiomas and acoustic neuromas are not seen well with CT. MRI is superior to CT in evaluating most cerebral parenchymal lesions. However, since CT requires less time than MRI, it may be the method of choice for some patients. Further, MRI cannot be used with patients who have cardiac pacemakers.

Magnetic Resonance Imaging (MRI). This methodology is based on the responses of protons in a strong magnetic field to the effect of a brief radiofrequency pulse. As described by S. Gilman (University of Michigan Medical Center), "Within the magnetic field, protons of the body align the axis of their nuclear spins with the external magnetic field and are displaced from this alignment by the radiofrequency pulse. When the pulse ends, the protons return to their previous orientation in the magnetic field, and radiofrequency energy is emitted from the tissue. The energy is amplified, recorded, and transformed into a planar or three-dimensional image of the structure.

Positron-Emission Tomography (PET). Known and used for several years, PET measures gamma rays emitted by radioactive tracer molecules, variously labled, incuding carbon-11, fluorine-18, and oxygen-15. The method has been useful for clinical diagnosis, but has faced numerous obstacles in developing a reputation as a research tool. PET does not necessarily yield images that correlate directly with brain activity. Although PET scanning can measure blood flow, some investigators observe that blood flow does not always represent an increase in neural activity. Others say that the method simply isn't fast enough. Marcus Raichle (Washington University Medical Center) has noted that investigators who got on the PET bandwagon in the early 1980s were not always the most rigorous of investigators, and one can become discouraged when "bad science" is carried out with PET. Nevertheless, improvements are continuing and Raichie observes, "The key issue is to introduce rigor into these experiments, such that people not directly involved will understand that this is a legitimate scientific tool, that this isn't just playing or picture taking." Researchers have developed new ways to analyze PET data, including the standardization of results by labeling parts of the brain with coordinates rather than names and using coordinates that provide the flexibility required for different human brain sizes studied."

The most successful results with PET to date have involved fairly simple systems, such as those of the visual and motor cortex, where there is already a good, basic understanding of the physiology.

Most researchers agree that knowledge of the human brain must stem from human brain studies and not extrapolated from patterns exhibited by other species. Admittedly, a comprehension of how the brain processes the problems of language and the handling of deep emotion remain to be learned a number of years hence.

Single-Photon-Emission Computed Tomography (SPECT). This methodology makes up for a major shortcoming of PET, namely that SPECT is not required to be near a source of cyclotron radiation. With SPECT, radionuclides that emit gamma rays are administered by intravenous injection or inhalation. These radionuclides have longer half-lives than positron-emitting radionuclides and thus commercially available substances can be used. Although SPECT is less costly than PET, the spatial resolution of the images is more limited. In recent times, SPECT has been used widely to study cerebral blood flow, cerebral blood volume, and cerebral neurotransmitter receptors. Abnormalities in cerebral blood flow can be demonstrated, as found in stroke, migraine, epilepsy, and degenerative diseases, such as Alzheimer's, Pick's, and Huntington's disease.

Brain Mapping Network. In late 1990, neuroscientists launched a computerized database to incorporate the tremendous amounts of data that describe the structure and locations of functions in the brain, ranging from sites active in vision and speech to those that are engaged in anxiety and other emotions. Known as *BrainMap*, the project was conceived by Peter Fox and colleagues at Johns Hopkins University, who observe, "We hope to develop a system similar to the genome database." It is planned to present information on the network as numerical coordinates that describe the precise locations of functions or tasks in the brain. One objective of the database is that of impelling neuroscientists to describe their data in uniform terms.

Computational neuroscience, commenced in a serious way in the 1980s, is gaining increased attention in the early 1990s. The objective of computational neuroscience is to delineate the manner in which electrical and chemical signals are used in the brain as a means for representing and processing information. At an increasing rate, computer-created brain models are used to relate the microscopic level through the use of molecular and cellular techniques with the systems level as defined through the study of behavior. Some questions posed to computational neuroscience include: (1) What kind of computer is the brain? (2) How can realistic brain models be created, and what is the best technology for brain modeling? (3) How does the brain detect and compute motion? (4) What are the principles of orientation selectivity in the visual cortex?

Sejnowski, Koch, and Churchland, researchers in this field, summarize a comprehensive paper on computational neuroscience with the observation: "At this stage in our understanding of the brain, it may be fruitful to concentrate on models that suggest new and promising lines of experimentation, at all levels of organization. In this spirit, a model should be considered a provisional framework for organizing possible ways of thinking about the nervous system." See reference listed.

The construction of mental models by readers or listeners of narrative material have been the center of recent studies by G. H. Bower and D. G. Morrow (Stanford University and Stanford Medical Center). The skill represented in building such models is, in essence, the basis of language comprehension. As observed by these researchers, "Cognitive psychologists and education specialists focus on research in reading comprehension, because it involves many components of intelligence, recognition of words, decoding them into meanings, combining meanings into statements, inferring connections among statements, holding in short-term memory earlier concepts while processing later discourse, inferring the writer's or speaker's intentions, schematization of the gist of a passage, and memory retrieval in answering questions about the passage. Thus, the study of comprehension has become for cognitive psychologists what the fruit fly became for geneticists, a means of investigating many issues."

In their paper (see reference listed), the researchers conclude, "The ability to construct and manipulate valid models of reality provides humans with our distinctive adaptive advantage; it must be considered one of the crowning achievements of the human intellect."

Neural Networks. In a review of neuroscience models and networks, M. Barinaga reviews the field, which dates back to the 1950s and 1960s. A breakthrough occurred in 1993 when T. Sejnowski (Johns Hopkins University) and G. Hinton (Carnegie-Mellon University) introduced the three-layer neural net. This proved to be a striking improvement over its predecessors. Barinaga observes, "In the three-layer net a middle layer of units connects the input and output layers. When the net is given an input, it sends signals through the hidden layer to produce an output. That output is checked against the "correct" output, and a learning algorithm is used to reduce error by strengthening or weakening connections in the net. The hidden layer retains the record of how the network distributed the task in the process."

Like other approaches, the three-layer network has its enthusiasts and its protagonists. With the appearance of more physiologically based models, stronger, more reliable predictions of brain activity can be expected. Most neuroscientists are looking to the current decade for large improvements in this science/art.

Neuronal Cell Development for Research

In years past, the number of nerve cells available for biochemical and physiological experiments has been limited by the comparatively small

TABLE 2. DISEASES AND DISORDERS OF THE NERVOUS SYSTEM AND BRAIN

Muscle And Neuromuscular Junction	Degenerative And Hereditary	Metabolic And Nutritional Disorders
Muscular dystrophy	Diseases	Encephalopathy
Toxic-metabolic myopathies	Alzheimer's disease	Wernicke's encephalopathy
Glycogen storage disease (McArdle's disease)	Arteriosclerotic dementia	Infections
Myoglobinuria	Occult hydrocephalus	Acute pyogenic meningitis
Malignant hyperthermia	Diseases of the basal ganglia	Chronic meningitis
Alcoholic myopathy	Huntington's disease	Acute amebic meningitis and
Drug-induced myopathy	Essential tremor	meningoencephalitis
Endocrine myopathy	Dystonia	Parameningeal infections
Inflammatory muscle disease	Shy-Drager syndrome	Demyelinating Disease
Polymyositis and dermatomyositis	Progressive supranuclear palsy	Multiple sclerosis
Myasthenia gravis	Wilson's disease	Cerebrovascular Diseases
Myasthenic (Eaton-Lambert) syndrome	Gilles de la Tourette's	Cerebral transient ischemic attacks
Peripheral Nervous System	syndrome	Ischemic stroke
Cranial mononeuropathy	Amyotrophic lateral sclerosis	Cerebral hemorrhage (hemorrhagic
Bell's palsy	Hereditary ataxias	stroke)
Trigeminal neuralgia (tic douloureux)	Spinocerebellar degeneration	Subarachnoid hemorrhage
Vestibular neuronitis	Joseph's disease	Hypertensive encephalopathy
Localized neuropathies	Neurocutaneous disorders	Cerebral arteritis
Carpal tunnel syndrome	Stiff-man syndrome	Headache
Ulnar palsy	Mechanical And Traumatic	Classic migraine
Brachial plexus neuropathy	Disorders	Common migraine
Meralgia paresthetica	Spinal cord disorders	Basilar artery migraine
Peroneal palsy	Cervical disk disease	Ophthalmoplegic migraine
Toxic or metabolic neuropathy	Lumbar disk disease	Hemiplegic migraine
Drug-induced	Claudication of the cauda equina	Cluster headache
Chemical and heavy-metal induced	Cervical fracture-dislocation	Increased Intracranial pressure
Vitamin B_{12} deficiency	Cervical spondylosis	Epilepsy
Developmental Anomalies	Head trauma	Sleep Disorders
Aqueductal stenosis	Neoplastic Disorders	Insomnia
Arnold-Chiari malformation	Brain tumor	Narcolepsy
Syringomyelia	Spinal cord tumor	Sleep apnea syndromes

NOTE: Several of these disorders are described in separate articles in this encyclopedia, or as parts of broader articles. Check alphabetical index.

numbers of cells that can be obtained from an animal. A cell line for searchers has been difficult to achieve because mature neurons do not divide and will resist manipulation to a continuously dividing state. Several laboratories recently have developed a neuronal cell line by manipulating neurons in an early and vulnerable stage of development—that is, before they have matured.

Investigators at the California Institute of Technology have developed a method for purifying cells and "immortalizing" them with special retroviruses. In this process, several new and important findings were made. The cells do not respond to nerve growth factor (NGF) until they are first exposed to the fibroblast growth factor (FGF). The latter causes the cells to create NGF receptors. The influence of NGF permits their conversion to sympathetic neurons. When infant neurons are not influenced by FGF, they move to the adrenal gland where they become secretory cells.

Brain/Nervous System Disorders

Major abnormalities resulting in diseases and disorders of the brain and nervous system are described elsewhere in this encyclopedia. Consult alphabetical index. An abridged list of these diseases is given in Table 2.

Additional Reading

Alkon, D. L.: "Memory Storage and Neural Systems," *Sci. Amer.*, 42–50 (July 1989).

Aminoff, M. J., Editor: "Neurology and General Medicine," Churchill Livingstone, New York, 1989.

Baddeley, A.: "Working Memory," *Science*, 556 (January 31, 1992).

Barinaga, M.: "Neuroscience Models the Brain," *Science*, 524 (February 2, 1990).

Barinaga, M.: "The Tide of Memory, Turning," *Science*, 1603 (June 29, 1990).

Barinaga, M.: "The Mind Revealed," *Science*, 816 (August 14, 1990).

Barinaga, M.: "Technical Advances Power Neuroscience," *Science*, 908 (November 16, 1990).

Barinaga, M.: "Ion Channel Research Wins Physiology Nobel," *Science*, 380 (October 18, 1991).

Barnes, D. M.: "NMDA Receptors Trigger Excitement," *Science*, 254 (January 15, 1989).

Bizzi, E., Mussa-Invaldi, A., and S. Giszter: "Computations Underlying the Execution of Movement: A Biological Perspective," *Science*, 287–291 (July 19, 1991).

Black, I. B.: "Information in the Brain. A Molecular Perspective," MIT Press, Cambridge, Massachusetts, 1991.

Bower, G. H., and D. G. Morrows: "Mental Models in Narration," *Science*, 44 (January 5, 1990).

Changeux, J-P, and M. Konishi, Editors: "The Neural and Molecular Bases of Learning," Wiley, New York, 1987.

Collingridge, G., Editor: "Neuropharmacology," Pergamon Press, Tarrytown, New York, 1993.

Crick, F., and C. Koch: "The Problem of Consciousness," *Sci. Amer.*, 153 (September 1992).

Davis, J. L., and R. W. Newburgh, Editors: "Brain Structure, Learning, and Memory," Westview Press, Boulder, Colorado, 1989.

Damasio, A. R., and H. Damasio: "Brain and Language," *Sci. Amer.*, 89 (September 1992).

Desimone, R.: "The Physiology of Memory: Recordings of Things Past," *Science*, 245 (October 9, 1992).

Edelman, G. M., Gall, W. E., and W. M. Cowan, Editors: "Synaptic Function," Wiley, New York, 1987.

Edelman, G. M.: "The Remembered Present: A Biological Theory of Consciousness," Basic Books, New York, 1989.

Efron, R.: "The Decline and Fall of Hemispheric Specialization," Erlbaum, Hillsdale, New Jersey, 1990.

Fischbach, G. D.: "Mind and the Brain," *Sci. Amer.*, 48 (September 1992).

Fuster, J. M.: "The Prefrontal Cortex," 2nd Edition, Raven, New York, 1988.

Gibbons, A.: "New Maps of the Human Brain," *Science*, 122 (July 13, 1990).

Gibbons, A.: "Did Cooler Heads Prevail?" *Science*, 1338 (December 7, 1990).

Gilman, S.: "Advances in Neurology," *New Eng. J. Med.*, Part I, 1608 (June 11, 1992); Part II, 1671 (June 18, 1992).

Gilroy, J.: "Basic Neurology," 2nd Edition, Pergamon Press, New York, 1990.

Goldman-Rakie, P. S.: "Working Memory and the Mind," *Sci. Amer.*, 111 (September 1992).

Glickstein, M.: "The Discovery of the Visual Cortex," *Sci. Amer.*, 118 (September 1988).

Gottlieb, D. I.: "GABAergic Neurons," *Sci. Amer.*, 82 (February 1988).

Gray, C. M., Konig, P., Engel, A. K., and W. Singer: "Oscillatory Responses in Cat Visual Cortex Exhibit Inter-columnar Synchronization Which Reflects Global Stimulus Properties," *Nature*, 334m Col. 383 (1989).

Gunderson, C.H.: "Essentials of Clinical Neurology," Raven Press, New York, 1990.

Harvey, P. H., and J. R. Krebs: "Comparing Brains," *Science*, 140 (July 13, 1990).

Hinton, G. E.: "How Neural Networks Learn from Experience," *Sci. Amer.*, 145 (September 1992).

Hoffman, M.: "A New Role for Gases: Neurotransmission," *Science*, 1788 (June 28, 1991).

Holden, C.: "What's On Your Mind? Check *BrainMap*." *Science*, 1203 (November 30, 1990).

Holden, C.: "New Center to Study Therapies and Ethnicity," *Science*, 748 (February 15, 1991).

Holloway, M.: "Under Construction: Temporary Scaffolding Guides Nerves in the Developing Brain," *Sci. Amer.*, 25 (January 1992).

Hubel, David H.: "The Brain" (September 1979). A Classic reference in *The Laureates' Anthology*, 26–35, Scientific American, Inc., New York, 1990.

Kalin, N. H.: "The Neurobiology of Fear," *Sci. Amer.*, 94 (May 1993).

Kandel, E. R., Schwartz, J. H., and T. M. Jessell, Editors: "Principles of Neural Science," 3rd Edition, Elsevier, New York, 1991.

Kandel, E. R., and R. D. Hawkins: "The Biological Basis of Learning and Individuality," *Sci. Amer.*, 79 (September 1992).

Kelner, K. L., and D. E. Koshland, Jr.: "Molecules to Models: Advances in Neuroscience," AAAS Books, Waldorf, Maryland, 1989.

Kien, J., McCrohan, C. R., and W. Winlow: "Neurobiology of Motor Programme Selection," Pergamon Press, Tarrytown, New York, 1993.

Kimbelberg, H. K., and M. D. Norenberg: "Astrocytes," *Sci. Amer.*, 66 (April 1989).

Kimura, D.: "Sex Differences in the Brain," *Sci. Amer.*, 119 (September 1992).

Kinoshita, J.: "Memory in a Neuron," *Sci. Amer.*, 28 (January 1989).

Kolb, B., and I. Q. Whislaw: "Fundamentals of Human Neuropsychology," 3rd Edition, Freeman, Salt Lake City, 1989.

Levi-Montalcini, R., and P. Calissano: "The Nerve-Growth Factor" (June 1979). A Classic reference in *The Laureates' Anthology*, 50–59, Scientific American, New York, 1990.

Lewin, R.: "Living in the Fast Track Makes for Small Brains," *Science*, 513 (October 28, 1988).

Llinas, R. R.: "The Intrinsic Electrophysiological Properties of Mammalian Neurons: Insights Into Central Nervous System," *Science*, 1654 (December 12, 1988).

Marsden, C. D., and T. J. Fowler, Editors: "Clinical Neurology," Raven Press, New York, 1989.

Marx, J.: "A New Link in the Brain's Defenses," *Science*, 1278 (May 29, 1992).

Miller, C.: "1990: Annus Mirabilis of Potassium Channels," *Science*, 1092–1096 (May 24, 1991).

Milner, P. M.: "The Mind and Donald O. Hebb," *Sci. Amer.*, 124 (January 1993).

Neher, E., and B. Sakmann: "The Patch Clamp Technique," *Sci. Amer.*, 44 (March 1992).

Nicoll, R. A.: "The Coupling of Neurotransmitter Receptors to Ion Channels in the Brain," *Science*, *241*, 545–551 (1988).

Norenberg, M. D., Hertz, L., and A. Schousboe: "Biochemical Pathology of Astrocytes," Alan R. Liss, 1988.

Paximos, G., Editor: "The Human Nervous System," Academic Press, San Diego, California, 1990.

Penner, R., and E. Neher: "The Patch-Clamp Technique in the Study of Secretion," *Trends in Neurosciences*, 159, Vol. 12, No. 4, (April 1989).

Purvis D.: "Body and Brain," Harvard Univ. Press, Cambridge, Massachusetts, 1988.

Sakmann, B.: "Elementary Steps in Synaptic Transmission Revealed by Currents Through Single Ion Channels" (Nobel Lecture). Elsevier Publishing Company, Amsterdam and New York, 1992.

Sakmann, B.: "Elementary Steps in Synaptic Transmission Revealed by Currents Through Single Ion Channels," *Science*, 503 (April 24, 1992).

Schlossberg, D., Editor: "Infections of the Nervous System," Springer-Verlag, New York, 1990.

Sejnowski, T. J., Koch, C., and P. S. Churchland: "Computational Neuroscience," *Science*, 1299 (December 9, 1988).

Selkoe, D. J.: "Aging Brain, Aging Mind," *Sci. Amer.*, 135 (September 1992).

Shatz, C. J.: "The Developing Brain," *Sci. Amer.*, 61 (September 1992).

Skerrett, P. J.: "Rewired Brains," *Technology Review (MIT)*, 15 (January 1991).

Spector, R., and C. E. Jlhanson: "The Mammalian Choroid Plexus," *Sci. Amer.*, 68 (November 1989).

Squire, L. R., and S. Zola-Morgan: "The Medial Temporal Lobe Memory System," *Science*, 1380 (September 20, 1991).

Stipp, D.: "Grand Gamble on Fruit Fly Learning," *Science*, 1486 (September 27, 1991).

Stuhmer, W.: "Structure-Function Studies of Voltage-Gated Ion Channels," *Neuron*, 665, Vol. 3, No. 6 (September 1989).

Sudarsky, L.: "Pathophysiology of the Nervous System," Little, Brown, Boston, Massachusetts, 1990.

Trevarthen, S., Editor: "Brain Circuits and Functions of the Mind," Cambridge Univ. Press, New York, 1990.
Tulving, E., and D. L. Schacter: "Priming and Human Memory Systems," *Science*, 301 (January 19, 1990).
Weiss, J.: "Unconscious Mental Functioning," *Sci. Amer.*, 103–109 (March 1990).
Wise, S. P., Editor: "Recent Explorations of the Brain's Emergent Properties," Wiley, New York, 1987.
Zeki, S.: "The Visual Image in Mind and Brain," *Sci. Amer.*, 69 (September 1992).
Zornetzer, S. F., Davis, J. I., and C. Lau: "An Introduction to Neural and Electronic Networks," Academic Press, San Diego, California (1990).

NERVOUS SYSTEM (Fishes). See **Fishes.**

NET. A set of intervals such that every point of a closed linear interval $[a, b]$ is contained in at least one interval of the set, each interval being called a *mesh* of the net.

NETTLE HAIR. A form of hairlike scale found on some caterpillars which causes an irritation of human skin resembling that of nettles. Caterpillars of the io and the brown-tail moths have larger poisonous spines of similar properties. The irritation caused by some species is very mild but others are much more severe.

NETWORK. In elementary terms, a network consists of three or more points that are connected physically by some medium for the purpose of transferring information, energy, or materials from one point to another. Networks may occur naturally, as in the case of the brain and nervous system of living creatures, or networks may be contrived to serve human needs. *Material-transfer* networks are exemplified by gas, oil, and water pipelines. An *energy-transferring* network is typified by the distribution over wires and cables of centrally generated electricity. An *information-transfer* network relies on various segments of the electromagnetic spectrum to convey intelligence from transmitters to two or more receivers. A simple connection between two points usually is referred to as a *link* rather than a network. As of the early 1990s, the word *network* most frequently refers to a *communication* network, the subject addressed by this article.

The points on a network may be referred to as stations or nodes at which information is received/transmitted instrumentally or humanly for storage, action, or sometimes retransmittal. Ultimately, unless wasted, the information will become part of the human need for that information. Information transmitted and received may be of analog or digital form, but the latter predominates by far. Thus, the term *digital network* is very common.

Networking equipment is costly, and the total cost of some nationwide and global networks is tremendous. Networks must be planned for expansion because of the need for more and more information exchange. The concept of information networks as they are known today essentially commenced with telegraphy and telephony. For decades, these were person-to-person communication. See **Telephony.**

Chronology of Computer Networks. The concept of a computer network in the United States and worldwide stemmed from research work at the Rand Corporation in the early 1960s. Late in the 1960s, a network sponsored by the U.S. Defense Advanced Research Projects Agency (DARPA) and known as ARPANET was established. According to D. Van Honweling (University of Michigan), "The notion when ARPANET was established was that it was primarily to share computing resources. As things turned out, that wasn't the way it got used. It got used by human beings who wanted to work with other human beings." In the mid-1970s, a variety of networks joined ARPANET to offer connectivity. These subsidiary networks had the specific objective of linking government and academic institutions together. Concurrently, NASA and the U.S. Department of Energy established their own networks.

Taking note of the lack of a national framework for networking as of the early 1980s, the National Science Foundation (NSF) established the NSFNET designed around the existing six NSF-supported supercomputer centers. The concept was that of including not only the supercomputer centers, but also to link together the several other important networks that had been developed during the mid-1980s. Operation of the network was assigned to private contractors when this "backbone" operated at 56 kilobits per second. A rate of 1.5 million bits per second had been achieved by early 1990, with goals as of 1993 set in excess of 50 million bits per second.

A *packet* is defined as bits that contain address and some fraction of the particular message being sent. On this standard, over 100 million packets per month had been sent in 1988. By 1990, this figure had increased to 2.5 billion packets. By February 1990, 10% of all information ever sent on the network had been sent during 1 month. A survey of NSFNET use indicates: (1) networked mail, 30%, (2) file exchange, 29%, (3) interactive applications, 20%, (4) name lookup, 15%, and (5) other services, 6%.

Of growing value to users of computer networks is the transmission of *motion graphics*. Most scientists agree that moving pictures are more eloquent than words. One scientist has observed, "Most people's comprehension works around visualization. Networks are making that comprehension tremendously more transferable."

Several years ago, the concept of a "Global Village" was envisioned, wherein communications, including graphics—essentially as an extension to traditional radio and television—would permit persons living almost anywhere to participate globally in science, politics, the arts, and other information exchange. Scientists who work on a single project, but at different locations, use networks to exchange concepts and ideas, problems and solutions, thus conserving on the expenses of establishing single quarters or of traveling a lot. But, from the shadows, the vision of computer hackers is seldom lost. Means to prevent system tampering must continue to be high on the list of designing future networked systems. One scientist has observed, "We really need to develop a set of cultural and behavioral protocols for using the networks!"

Data Transmission Needs Are Virtually Unlimited. Considering the information worthy of recording that has been generated in centuries past and the macro amounts of data generated each second of every day in present times that are important to some segment of civilized society, the word *overwhelming* seems inadequate. Thousands of needs to retrieve data can be cited, of which just one case—that is, the needs of the National Aeronautics and Space Administration (NASA)—are exemplary. Considering just the time frame that embraced NASA data gathering from spacecraft (*Pioneer* to the *Ultraviolet Explorer*), 6 trillion bytes (1 byte = 8 bits or binary units) were collected. Spacecraft launched during the last few years have been estimated to more than quadruple that amount. Although currently not fully operable, the *Hubble Space Telescope* was estimated to generate several trillion bytes each year. Introduction of the *Earth Observing System* (EOS) is estimated to generate a trillion bytes of information every few days. (The battery of sensors planned for EOS have been so designed.) Because of data storage and retrieval costs, these are a prerequisite cost accounting factor in budgeting for total mission coverage and, in fact, for selecting only those missions for which adequate data handling support will be available the years that the mission may be productive.

By way of comparison, it has been estimated that the U.S. Library of Congress, with over 19 million books, represents the storage of 6 trillion bites of information.

The data storage and retrieval components of the human genome project also are exemplary of how failure to adequately budget for electronic storage and networking of monstrous quantities of information can threaten ambitious research projects per se. An accounting of the *Bionet* project (National Institutes of Health) is given in the Wallich reference listed.

Examples of network databases are those of the Scientific & Technical Information Network, including such topics as: (1) BEILSTEIN, a comprehensive database in organic chemistry offering a wide range of information for over 1,700,000 substances reported over the past 150 years, (2) JANAF, a joint U.S. Army-Navy-Air Force file produced by the National Institute for Standards and Technology (NIST), containing the chemical thermodynamic properties for over 1000 substances, (3) NBSFLUIDS, which is an NIST file that contains calculation and data generation software useful for creating tables of temperature-dependent thermodynamic and thermophysical properties for 12 cryogenic fluids, and (4) NBSTHERMO, an NIST file of ambient temperatures and chemical thermodynamic properties of 8000 inorganic and numerous two-carbon organic substances.

Expanding the Points in a Network. Even though most of the time communications and transportation networks usually are initially designed to accomodate for future as well as present needs, in recent years the satiation of network users is seldom accommodated. The frequent to requirement to expand networks today is regarded by many network engineers as virtually inevitable. What initial steps can be taken to assure a minimum cost for network expansion? As reported by B. Cipra (see reference listed), F. Hwang (AT&T Bell Laboratories) and Ding Zhu Du (Princeton University) have proposed the method indicated in Fig. 1. This materializes a conjecture with which mathematicians have been attempting to prove for many years. In formulating their approach, the engineers have borrowed from the "minimax" problem of game theory. See also **Game Theory.**

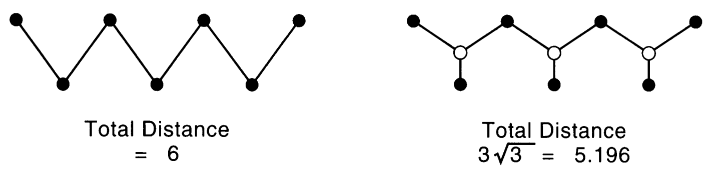


Fig. 1. Designing the "shortest" network. As shown here, a network can be shortened by adding additional points. The example shows a savings of about 13%. Whether or not savings of over 13% can be achieved has not been settled among mathematicians. As indicated by this figure, the greatest savings result when the points are at the vertices of an equilateral triangle. Using the center of the triangle as a hub reduces the length of the network by $\frac{\sqrt{3}}{2}$ (~13%).

The "shortest network" problem has been analyzed in scholarly detail by M. Bern (Xerox Palo Alto Research Center) and R. Graham (AT&T Bell Laboratories). In their paper (see reference listed), the researchers conclude: "Although knowledge about algorithms (including Melzak's) has progressed greatly in recent years, the shortest-network problem remains tantalizingly difficult. The problem can be stated in simple terms, and yet solutions defy analysis. A tiny variation in the geometry of a problem may appear to be insignificant, and yet it can radically alter the shortest network for the problem. The shortest-network problem will continue to frustrate and fascinate us for years to come."

Classification of Network Uses. To ask the question, "Where are communication networks used?" is tantamount to asking, "Where is information used?" An appreciation for the phenomenal needs and uses of networking may be piqued by considering the following:

- Statistical information for business and government. One of the early pioneering users of networks was the travel industry for determining space availability and making reservations. Computerized banking and trading also are early examples of early network use.
- Educational institutions, where most universities today have extensive private networks and the ability to tap into national and global networks. This is expanding rapidly to include public and vocational schools.
- Scientific research, notably for efforts conducted by teams of researchers who represent several disciplines and who do not have their offices and laboratories at a centralized location.
- Transportation and shipping, where goods movements must be closely time related. See also **Navigation.**
- Medical practice, particularly pertaining to crises that must tap distant sources of special expertise and in order to expedite the location and ultimate receipt of organs and fluids for transplantation. Numerous hospitals operate their own networks, which can readily be tapped into national and global networks.
- Earth and space research, as previously mentioned.
- Entertainment program retrieval.
- Manufacturing production, described later in this article.
- Distribution systems, to which much attention has been devoted by economists recently.
- And this list could go on and on.

Distribution and Networking. Information networking has and will continue to impact upon the very fundamental life-styles of business, government, scientific, and other professional people. Quite recently, particular attention has been given to the manner in which networking ultimately may affect materials distribution, selling, and trading. Depending upon a nation's form of government, prices may be fixed or tightly regulated by a central organization or allowed to float in accordance with supply and demand. Even in free-trade economies, some commodity prices (notably of utilities, such as the gas and electricity supply) are regulated by national or state and provincial authorities. As observed by K. McCabe, S. Rassenti, and V. L. Smith (University of Arizona), "Domestically, we have witnessed in the last decade (1980s) uncommon political and economic forces that have resulted in increased reliance on markets to discipline prices, output, and the entry and exit of firms in industries traditionally regulated by state and federal agencies. This has been part of a worldwide move toward privitization in the socialist and command economies of Great Britain, New Zealand, Eastern Europe, and the (former) Soviet Union. In the United States the extent of deregulation has been less than complete in all of them."

The deregulation movement motivated an experimental study of *auction markets* designed for interdependent network industries, such as natural gas pipelines or electric power systems by the aforementioned researchers. Through the use of an information-transfer network and a computerized dispatch center, decentralized agents would submit bids to buy commodities and offers to sell transportation and commodities. Computer algorithms would determine prices and allocations that maximize the gains from exchange in the system relative to the submitted bids and offers. Details are described in the McCabe reference listed.

Networked Communications for Manufacturing

Although information-exchange networks for manufacturing and processing facilities frequently include wide-area networks (WANs), which exchange information between plants at different locations and with corporate headquarters, the principal need for information exchange occurs at the factory floor level through the use of *local-area networks* (LANs), which provide the coordination and govern the actions of specific machines and processes. See also article on *Local Area Networks.*

As described in the article on **Control System**, industrial manufacturing is based largely on the measurement and control of numerous variables (temperature, pressure, dimension, count, material flow, to mention a few). In a typical manufacturing scene, hundreds to thousands of feedback control loops may be present. Information coordination and balance are achieved by way of local area networks.

Because of the multitude of connections required and the great variety of sensors used, severe standards for component design are imposed. Development of these standards is described in the article on Control System Architecture.

Other Networks

The word *network* is used to designate systems other than information-exchange networks. For example, *electric network* stems back to Kirchhoff (1850) and Maxwell (1870).

An electric network may be described as a combination of elements, either as (1) a combination of interconnected devices, such as inductors, capacitors, resistors, and generators, or (2) the abstraction of interconnected branches having the properties of inductance, resistance, capacitance, etc.

An *active network* is one whose output waves are dependent upon sources of power, apart from that supplied by any of the actuating waves, which power is controlled by one or more of these waves.

An *all-pass network* is designed to introduce phase shift or delay without introducing appreciable attenuation at any frequency.

A *differentiating network* is one whose output is the time derivative of its input waveform. Such a network preceding a frequency modulator makes the combination a phase modulator; or, following a phase detector, it makes the combination a frequency detector. Its ratio of output amplitude to input amplitude is proportional to frequency, and its output phase leads its input by 90 degrees.

An *equivalent network* is one in which, under certain conditions of use, it may replace another network. The networks need not be of the same form. For example, one may be electrical, the other mechanical. If one network can replace another network in any system whatsoever without altering in any way the operation of that portion of the system external to the networks, the networks are said to be of *general equivalence*.

A *linear-passive network* is a network such that (1) if currents of any waveform are fed to the terminals of the network, the total energy delivered to the network is non-negative; (2) no voltages appear between any pair of terminals before a current is fed to the network.

Neural networks are described in the article on **Nervous System and the Brain**.

Additional Reading

Bern, M. B., and R. L. Graham: "The Shortest-Network Problem," *Sci. Amer.*, 84 (January 1989).
Cerf, V. G.: "Networks," *Sci. Amer.*, 72 (September 1991).
Cipra, B.: "In Math, Less is More—Up to a Point," *Science* 1081 (November 23, 1990).
Dertuozos, M. L.: "Building the Information Marketplace," *Technology Review (MIT)*, 29 (January 1991).
Haavind, R.: "The Smart Tool for Information Overload—*Hypertext*," *Technology Review (MIT)*, 42 (November–December 1990).
Horgan, J.: "Jukeboxes for Scientists," *Sci. Amer.*, 24 (July 1989).
Kauffels, J.: "Practical LANs Analyzed," Halstead Press, New York, 1989.
Kay, A. C.: "Computers, Networks and Education," *Sci. Amer.*, 138 (September 1991).
King, J. P.: "Distributed Control Systems," 3.5 in *Process/Industrial Instruments and Controls Handbook*, 4th Edition (D. M. Considine, Editor), McGraw-Hill, New York, 1993.
Malone, T. W., and J. F. Rockart: "Computers, Networks and the Corporation," *Sci. Amer.*, 128 (September 1991).
McCabe, K. A., Rassenti, J., and V. I. Smith: "Smart Computer-Assisted Markets," *Science*, 534 (October 25, 1991).
Negroponte, N. P.: "Products and Services for Computer Networks," *Sci. Amer.*, 106 (September 1991).
Palca, J.: "Getting Together Bit by Bit," *Science*, 160 (April 13, 1990).
Peterson, I.: "The Electronic Grapevine," *Science News*, 90 (August 11, 1990).
Tesler, L. G.: "Networked Computing in the 1990s." *Sci. Amer.*, 86 (September 1991).
Wallich, P.: "Who's Minding the Store," *Sci. Amer.*, 20 (October 1989).

NETWORK SYNTHESIS. Network synthesis is that branch of the theory of electric networks which deals with the systematic determination of the structure and element values of an electric network possessing preassigned characteristics.

NETWORK (Telecommunications). See **Telephony.**

NEURALGIA. Pain occurring along the course of a cranial or peripheral nerve or posterior root. Neuralgia is difficult to differentiate sharply from neuritis. The primary distinction is that neuritis best describes the acute and chronic inflammation of the peripheral nerves, while neuralgia is due to acute and chronic inflammation of the pathway stations (ganglia) lying along the course of nervous pathways. The pain in neuralgia is usually of a sharp, shooting, intermittent type and accompanied by increased sensitivity of the skin supplied by the nerve. Many varieties of neuralgia are differentiated according to the body part affected. Common forms include: (1) trigeminal neuralgia, a very severe form marked by agonizing pain over branches of the trigeminal nerve in the face, (2) intercostal neuralgia, which can be mistaken for pleurisy because the intercostal nerves, after leaving the spinal cord, run around each side of the chest between the ribs, (3) Morton's neuralgia—pain in the joint of the third and fourth toe caused by pinching of a nerve in this region, and (4) sciatic neuralgia.

Trigeminal neuralgia also is known as tic douloureux and involves one or more branches of the fifth cranial (trigeminal) nerve. A closely associated neuralgia involves the ninth cranial nerve, often causing severely sharp pains in the throat, spreading to the ear on the same side of the face. Palliative relief in some cases can be obtained from injections of alcohol into the ganglion of the nerve. In more severe cases, surgical procedures, as cutting the sensory root of the nerve, may be required to provide permanent relief. Drugs such as *Dilantin*® Sodium slow the rate of peripheral nerve conduction and have been used in the treatment of trigeminal neuralgia. The successful use of injections of vitamin B_{12} also has been reported. Intercostal neuralgia is sometimes accompanied by a skin eruption in herpes zoster or shingles.

NEURITIS. An inflammation of a nerve, either chronic or acute. Mononeuritis or localized neuritis is the term used when one nerve is involved. When several or many nerves are involved the term multiple neuritis or polyneuritis is used. Localized neuritis develops from injury, infection, chronic intoxication, or metal poisoning. Neuritis, either localized or multiple, complicates many of the infectious diseases, such as typhoid, diphtheria, tuberculosis, smallpox, etc. Other causes include pressure on a nerve by a tumor or calcium deposits in osteo-arthritis, etc.

Multiple or polyneuritis is inflammation and degeneration affecting the peripheral nerves, usually in their distal portions. The commonest cause is dietary deficiency, primarily of vitamin B_1 (thiamine) which is usually associated with chronic alcoholism; other causes are various toxins, either chemical agents such as arsenical drugs, lead, nitrobenzol, coal-tar products, or toxins produced by bacteria.

The clinical picture is similar no matter what the cause: there is a gradual onset of numbness and tingling of the hands and feet, weakness of the limbs, altered sensation to touch, and pain, loss of deep reflexes, and eventually paralysis and atrophy of the extremities. Treatment is directed first toward removal of the cause.

NEUROFIBROMATOSIS. A condition characterized by multiple tumors in the skin or along the course of peripheral nerves. These tumors occassionally become malignant. Also called von Recklinghausen's disease.

NEUROLETPIC MALIGNANT SYNDROME. First described and named by Delay and Deniker in 1968, neuroleptic malignant syndrome (NMS) was thought to be a variant of drug fever in which hyperpyrexia and autonomic and other neurologic abnormalities developed during phenothiazine therapy. The principal characteristics of NMS are hyperthermia, hypertonicity of skeletal muscles, and fluctuating consciousness, along with instability of the autonomic nervous system. As described by Guzé and Baxter (see reference), common autonomic dysfunctions include pallor, diaphoresis, blood-pressure instability, tachycardia, and cardiac dysrhythmias. The muscular hypertonia consists of a generalized "lead-pipe" increase in tone, which increased muscle tone may result in decreased chest-wall compliance and, as a consequence, breathing problems severe enough to require respiratory support. Disturbances of consciousness may range from agitation or alert mutism to stupor and coma. Estimates of mortality range from 20 to 30%, this depending upon the type of neuroleptic drugs responsible and upon duration of exposure to offending drug. Deaths usually occur between 3 and 30 days after presentation of symptoms. Death may result from respiratory failure, cardiovascular collapse, renal failure, arrhythmias, and thrombo-embolism. Respiratory failure may result from aspiration pneumonia or tachypneic hypoventilation. A primary factor in the causation of NMS may be a decrease in the availability of dopamine in the brain.

Neuroleptic drugs, also referred to as antipsychotic agents or major tranquilizers have been commonly used in the treatment of psychotic illnesses and also used in general medicine as antiemetics; in connection with dissociative anesthesia; and for the treatment of diseases, such as Tourette's syndrome. Although neuroleptic agents have been known for side-effects, NMS, frequently fatal, has not been well recognized.

In 1985, the seriousness of the side effects of neuroleptic or antipsychotic drugs, such as phenothiazines, butyrophenones, and thioxanthenes, among others, prompted the American Psychiatric Association to issue a warning to psychiatrists and physicians, advising them to carefully weigh the benefits of the drugs against the potential for developing *tardive dyskinesia* (or NMS).

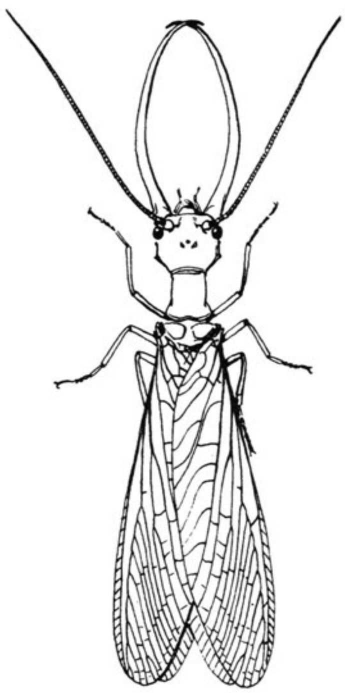
Additional Reading

Caroff, S. N.: "The Neuroleptic Malignant Syndrome," *J. Clin. Psychiatry*, **41**, 79–83 (1980).
Culliton, B. J.: "Antipsychotic Drugs (and Tardive Dyskinesia)," *Science*, **229**, 1248 (1985).
Diamond, J. M., and A. B. Santos: "Unusual Complications of Antipsychotic Drugs," *Am. Fam. Physician*, **26**, 153–157 (1982).
Guzé, B. H., and L. R. Baxter, Jr.: "Neuroleptic Malignant Syndrome," *N. Eng. J. Med.*, **313**(3), 163–165 (July 18, 1985).

Tollefson, G. I.: "The Neuroleptic Syndrome and Central Dopamine Metabolites," *J. Clin. Psychopharmacology*, **4**, 150–153 (1984)..

NEURON. See **Nervous System and the Brain.**

NEUROPTERA. Insects of varied form and habits, including the dobson fly, the golden eyes or lacewings, the alder flies, and the ant lions. The order is characterized by the four membranous wings, usually with a large number of branching veins which are united in some species to form a network. The mouth is formed for biting but in some larvae the mandibles and maxillae fit together to form a piercing and sucking organ. The insects are predacious.



Dobson fly, valued as a food for fishes. (*USDA*.)

Neuroptera are of minor economic importance as food for fishes. The larvae of lacewings prey on aphids, hence they are of some assistance in holding these pests in check.

The order contains about 1700 species. The main families of Neuroptera include:

Chrysopidae	Lacewings, aphid-lions, golden-eyed flies
Hemerobiidae	Brown lacewings
Mantispidae	Mantid flies
Myrmeleonidae	Ant lions or doodle bugs
Rhaphidiidae	Snake flies
Sialidae	Dobson flies, alder flies, fish flies

NEUROSYPHILIS. See **Syphilis.**

NEUROTENSION. Nervous System and the Brain.

NEUTRAL. 1. Having no electric charge, or no net electric charge. (Thus, an atom, in which the total negative charge of the electrons is equal to the positive charge of the nucleus which they surround, is neutral.)

2. According to the ionization hypothesis, a concentration of hydrogen ions equal to 1×10^{-7}. (The figure varies a little according to the temperature and the method of determining the degree of ionization of water.) Hydrogen-ion concentrations greater than this figure confer acid properties; lower concentrations occur in alkaline systems.

NEUTRALIZATION. See **Acids and Bases.**

NEUTRINO. A neutral particle of very small (presumed zero) rest mass and of spin quantum number $\frac{1}{2}$. This particle was initially postulated to account for the continuous energy distribution of beta particles and to conserve angular momentum in the beta-decay process. Experimental evidence indicates that, for the linear momentum to be conserved in the beta process, there must be a contribution from a departing neutrino. Presumably, a neutrino (or antineutrino) is emitted in every beta transition. The energy of a neutrino emitted in a beta disintegration is assumed equal to the difference between the energy of the particular beta particle and the energy corresponding to the upper limit of the continuous spectrum for that beta transition. The neutrino has also been postulated as one of the particles in pion (π) decay and as two of the particles in muon (μ) decay. These processes, however, lead to two types of neutrinos: an electron-associated neutrino ν_e and a muon-associated neutrino ν_μ. For example, $\pi^+ \rightarrow \mu^+ \rightarrow \bar{\nu}_\mu + e^+ + \nu_e$ or $\pi^+ \rightarrow e^+ + \nu_e$, whereas neutron decay obeys only the process $n \rightarrow p^+ + e^- + \bar{\nu}_e$, where the bar over ν indicates an anti-particle. The difference between neutrinos was established at Brookhaven in 1962 when it was shown that a beam of neutrinos from the process $\pi^+ \rightarrow \mu^+ + \nu_\mu$ gave rise to the process $\nu_\mu + n \rightarrow p + \mu^-$ but not to $\nu_\mu + n \rightarrow p + e^-$. There is also a neutrino associated with the tau particle. Because of its properties, the neutrino has negligible interactions with matter and has proved difficult to detect. It was first positively identified experimentally in 1956 by Reines and Cowan, Jr. See also **Particles (Subatomic).**

The term antineutrino usually denotes an antiparticle whose emission is postulated to accompany radioactive decay by negatron emission, such as, for example, in neutron decay into a proton p^+, negatron e^- and antineutrino $\bar{\nu}_e$, expressed by the equation $n \rightarrow p^+ + e^- + \bar{\nu}_e$. Capture of a neutrino by the neutron, $\nu_e + n \rightarrow p^+ + e^-$ would be an equally good description of the process. Positron emission is accompanied by a neutrino, as in the decay ${}^{64}Cu \rightarrow {}^{64}Ni + e^+ + \nu_e$. Orbital electron capture also involves a neutrino, as for example, $e^- + {}^{64}Cu \rightarrow {}^{64}Ni + \nu_e$. Since there is no possibility of charge differentiation between the antineutrino and the neutrino, differentiation between these two particles can be made only on the basis of such properties as the sign of the ratio of magnetic moment to angular momentum.

In the past the terms neutrino and antineutrino were sometimes used in reverse sense to that stated above, i.e., the neutrino is said to accompany negatron emission and the antineutrino, positron emission. The preferred usage has been accepted in order to provide conservation of leptons in the conservation laws.

Neutrino Astronomy. For several years, there has been a marked trend in astronomy to expand the use of the electromagnetic spectrum beyond the visual range in investigating the universe. It now appears that, in addition to infrared, ultraviolet, gamma ray, x-ray, and radio astronomy, among others, increasing attention will be given to neutrino astronomy, i.e., the detection and measurement of neutrinos emanating from such celestial bodies as supernovas and x-ray double stars. Traditionally, neutrinos have been created in large accelerators and used for investigating the characteristics of other elementary particles. It is now reasoned that neutrinos entering the earth's atmosphere from celestial distances may furnish new, heretofore unavailable information.

Search for Neutrinos

Certain mysteries continue to surround the neutrino, emphasizing how far physicists still may be from a full comprehension of the complete array of subatomic particles. Numerous detectors in different locations have been constructed for detecting neutrinos, particularly as they emanate from the sun.

The first recorded attempt to construct a neutrino telescope was undertaken by Davis (Brookhaven National Laboratory) in the 1960s. Davis' principal objective was detection of low-energy neutrinos emitted by the sun. These neutrinos are generated deep within the sun as the result of thermonuclear reactions. It has been estimated that nearly 10% of the energy released by the transmutation of hydrogen in the sun is carried away by neutrinos, which have energies ranging from $\frac{1}{2}$ million eV to about 14 million eV. The solar flux of neutrinos is tremendous. It is estimated that 10^{14} solar neutrinos pass through the human body every second. Davis' detector consisted of a large tank containing 610 tons of tetrachloroethylene, C_2Cl_4. Of the chlorine atoms present, 25% are of the isotope chlorine 37. When this atom captures a neutrino, it is transformed to an atom of argon 37.

An early detector was established in the shaft of the Homestake Mine (South Dakota) in 1968. The data gathered did not correspond with the scientists' expectation. The neutron flux measured was less than one-third that predicted. This discrepancy challenged researchers to question perhaps the "established" concepts of particle physics and indeed the manner in which the sun functions.

Some years later, a Japanese-built detector (*Kamiokande II*), designed to detect the more energetic solar-emitted neutrinos (~5 mil electron volts), came up about 50% short of the expected counts, thus reconfirming a shortage of solar neutrons.

Neither of the aforementioned detectors was designed to sense comparatively low-energy (proton-proton) neutrinos, which result from the fusion of two protons.

Subsequently, two additional detectors were built with the objective of seeking lower-energy neutrinos. One of the detectors was located in the Caucasus of the former Soviet Bloc and was named the Soviet-American Gallium Experiment (SAGE). The detector used a gallium metal detector believed to be sensitive to neutrinos, with energy as low as 0.23 mil electron volts. Another similar detector (*Gallox*) was constructed in Italy. At both sites, difficulties with calibration were expected and did occur. Some tenuous observations later were made to suggest a correlation of neutrinos detected with the occurrence of sun spots and with solar acoustic oscillations.

Based upon the foregoing experiences, some researchers observed that the same reluctance to interact with matter is responsible for the neutrino's long range and ability to resist detection. Thus, it was reasoned that an apparatus for detecting neutrinos should be massive and shielded from the interference of other particles and radiation. As a solution to these problems, some researchers proposed a deep underwater muon and neutrino detector (acronym DUMAND).

Fiber optic data cable will stretch from the shore for 30 km to a connector box some 4800 meters below the ocean surface. Strings of nine separate cables will rise vertically about 280 meters above the ocean floor. Each cable, held up by a float, will contain 24 detectors. The apparatus, referred to as a neutrino telescope, will have to pick up at least twn muon events per year from any given 1° patch of sky for the DUMAND scientists to be confident that they have a significant neutrino source, not just a few background pulses from non-neutrino cosmic rays.

Construction of DUMAND, with Japanese and German collaborators, got underway off the west coast of Kona (Hawaii) in 1990 and is expected to be operational within a few years. V. Stenger (University of Hawaii) observes, "The idea of using the ocean floor as a detector actually goes back to the 1960s and people started taking it seriously back in the mid 1970s."

A number of scientists are hopeful that the neutrino telescope will ultimately yield much additional information on neutron stars, supernovas, quasars, etc., which are presumed to be large emitters of neutrinos. Some astronomers estimate that SS 433 puts out 1000 times more energy per second than the brightest stellar object known in the galaxy. Why should such a powerful object have been discovered so recently? The fact is that SS 433 is a comparatively weak source of photons. The accreting matter that gives SS 433 its great power also serves to screen its bright central region from view. Much remains to be learned concerning SS 433 and perhaps neutrino astronomy may supply some answers at a future date.

Laboratory Experimentation on Neutrinos

During the 1960s, L. M. Lederman, M. Schwartz, and J. Steinberger conducted the well-known two-neutrino experiment, which established a relationship between particles, muon and muon neutrinos, electron and electron neutrino. This later evolved into the standard model of particle physics. The Nobel prize in physics was shared by these researchers in 1988.

Massive Neutrino Proposed. J. Simpson (University of Guelph, Ontario) in the late 1980s presented evidence of the existence of a *heavy neutrino* having a mass of 17,000 electron volts (keV, the units of energy that are interchangeable with mass). A renowned neutrino physicist for several years prior, in 1985 Simpson conducted a series of experiments in his laboratory. The objective of his initial experiment was that of measuring the energy of electrons emitted from heavy hydrogen (tritium) in the radioactive process of beta decay. In this process, the energy of the emitted electrons should appear as a spectrum (smooth curve) from zero to a maximum endpoint. However, Simpson noted "occasional" aberrations in the plotted data. This so-called "kink" corresponded with an energy of 17 KeV of the normal plot. This indicated the probable presence of an unknown massive force. In addition to repeated experiments with tritium, Simpson also used Sulfur-35. Simpson then suggested that the aberration, when it occurred, could be attributed to a heavy 17 KeV neutrino.

Simpson's carefully prepared data, when published, attracted wide attention and drew a wide spectrum of professional reactions. Meanwhile, experiments by other laboratories have confirmed Simpson's results. Researchers at the Lawrence Berkeley Laboratory, using Carbon-14, have reported evidence for a 17.2 KeV neutrino in 1.4% of their experiments. Researchers at the Ruder Boskovic Institute in Zagreb (formerly Czechoslovakia), using Iron-55 and Germanium-71, also have found the evidence appearing in 1.5% of their experiments. Thus, confidence in Simpson's claim appears to be gaining momentum as of the early 1990s.

As pointed out by M. Turner (University of Chicago), a massive neutrino would "violate every theoretical prejudice we have in particle physics, astrophysics, and cosmology." J. Bahcall (Institute of Advanced Study at Princeton) observes, "It's a surprise. If it's true, then it's pointing us in a different direction than previous physics suggested."

See also **Particles (Subatomic.)**

Additional Reading

Breuker, H., et al.: "Tracking and Imaging Elementary Particles," *Sci. Amer.*, 58 (August 1991).

Brown, L. M., Dresden, M., and L. Hoddeson, Editors: "Pions to Quarks: Particle Physics in the 1950s," Cambridge University Press, New York, 1989.

Cence, R. J., et al., Eds.: "Neutrino 81," University of Hawaii, Honolulu. 1981.

Dar, A.: "Astrophysics and Cosmology Closing in on Neutrino Masses," *Science*, 1529 (December 14, 1990).

Dehmelt, H.: "Experiments on the Structure of an Individual Elementary Particle," *Science*, 539 (February 2, 1990).

Florini, E., Ed.: "Neutrino Physics and Astrophysics," Plenum, New York, 1982.

Gutbrod. H., and H. Stöcker: "The Nuclear Equation of State," *Sci. Amer.*, 58 (November 1991).

Horgan, J.: "Three Americans Honored for 1960's Neutrino Experiment," *Sci. Amer.*, 31 (December 1988).

Lederman, L. M.: "Observations in Particle Physics from Two Neutrinos to the Standard Model," *Science*, 664 (May 12, 1989).

Powell, C. S.: "Looking for Nothing: The Taciturn Neutrino Keeps Physicists Guessing," *Sci. Amer.*, 22 (April 1991).

Schwartz, M.: "The First High-Energy Neutrino Experiment," *Science*, 1445 (March 17, 1989).

Selvin, P.: "Is There a Massive Neutrino?" *Science*, 1426 (March 22, 1991).

Traweek, S.: "Beamtimes and Lifetimes. The World of High Energy Physics," Harvard University Press, Cambridge, Massachusetts, 1988.

Waldrop, M. M.: "A Nobel Prize for the Two-Neutrino Experiment," *Science*, 669 (November 4, 1988).

Waldrop, M. M.: "Solar Neutrino Deficit Confirmed?" *Science*, 1607 (June 29, 1990).

Waldrop, M. M.: "Astrophysics in the Abyss," *Science*, 208 (October 12, 1990).

NEUTRON. The discovery of the neutron by Chadwick in 1932 represented a great step forward in the investigation of nuclei of atoms. Chadwick found that a radiation emitted when α-rays from polonium reacted with beryllium could project protons from a thin sheet of paraffin wax. Although the radiation itself produced no observable ionization when passing through a gas, the protons released from the paraffin were detected in an ionization chamber. Inability to produce ionization was interpreted as a lack of electric charge. From measurements of the ionization from the protons, Chadwick deduced that the so-called beryllium radiation must consist of neutral particles with a mass very nearly equal to that of the proton. He announced the discovery of the neutron, a previously unknown particle. It has been confirmed that the neutron has no charge and a mass of 1.088665 atomic mass units. Thus, it is heavier than the proton by 0.00139 mass unit. The introduction of the neutron into nuclear structure produced a sharp change in previously held concepts. Lacking knowledge of the neutron, masses of atomic nuclei had been attributed solely to protons. The number of protons required on this basis for most nuclei greatly exceeded the known charge number. In an attempt to solve this dilemma, a number of electrons were assigned to each nucleus to adjust the charge number to the proper value. This compromise created an even greater problem, that of accommodating so many electrons in the small space occupied by a nucleus. Bringing the neutron into the picture meant that a nucleus contains only enough protons to equal the charge number, with the rest of the mass contributed by neutrons. No additional electrons were required.

Decay. The neutron in the free state undergoes radioactive decay. Elaborate experiments by Robson were required to identify the products of the decay and to measure the half-life of the neutron. He showed that the neutron emits a β-particle and becomes a proton. The half-life was found to be 12.8 minutes. In stable nuclei, neutrons are stable. In radioactive nuclei, decaying by β-emission, the neutrons decay with a half-life characteristic of the nuclei of which they are a part. See also **Radioactivity.**

Detection. Because it is a neutral particle the neutron is detected by means of a secondary charged particle which it releases in passing through matter or by means of the radioactivity which the neutron can induce in stable elements. Protons may be projected by collisions with neutrons in hydrogenous material and the ionization from the protons can be measured in an ionization chamber, as in the original experiment with neutrons. Secondary charged particles may be the direct result of nuclear disintegration produced by neutrons, as in the case of the reaction $^{10}B + {}^{1}n \rightarrow {}^{7}Li + \alpha$. Commonly, the radioactivity induced in a stable element by neutron capture serves to detect neutrons, and this technique is known as the *activated foil method.* Also, fission may be utilized for detection of neutrons by placing fissionable material inside an ionization chamber and observing the ionization generated by the fission fragments.

Energies. The kinetic energy of neutrons has an important bearing on their behavior when interacting with nuclei. These kinetic energies may range from near zero to as much as 50 MeV. It is therefore natural to classify neutrons in terms of energy according to their properties in each range of energy. For example, energies from zero to about 1000 eV are usually called *slow neutrons*. Because they are more readily captured by nuclei than faster neutrons, slow neutrons are responsible for a large number of nuclear transformations. When slow neutrons have velocities in equilibrium with the velocities of thermal agitation of the molecules of the medium in which they are situated, they are called *thermal neutrons*. The distribution of these velocities approaches the Maxwell distribution

$$dn\,(v) = Av^2 e^{-(Mv^2/2kT)} dv$$

where v is the neutron velocity, M its mass, k is Boltzmann's constant, and T is the absolute temperature. In the slow neutron range of energies, various atomic nuclei show strong absorption (capture) of neutrons at fairly well-defined energies. Neutrons having energies corresponding to those of the absorption bands are called *resonance neutrons*. Frequently, neutrons with energies greater than 1000 eV and less than 0.5 MeV are termed *intermediate neutrons*. In more general terms, all neutrons with energies greater than 0.5 MeV are called *fast neutrons*. The practical upper limit of neutron energy is set by the devices thus far developed for accelerating charged particles to extremely high energies.

Magnetic Moment and Spin. Alvarez and Bloch succeeded in measuring the moment of the magnetic dipole associated with the known spin of $\frac{1}{2}$ possessed by the neutron. More refined measurements by Cohen, Corngold, and Ramsey of the magnetic moment μn yielded a value of

$$\mu n = 1.913148 \text{ nuclear magnetons}$$

Interactions with Nuclei. Neutrons may be scattered or captured by heavy nuclei. Scattering may be elastic, resulting only in the change of direction of the neutrons, or inelastic, in which the neutron loses part of its energy to the scattering nucleus. Collisions with light nuclei, in the absence of capture, result in communicating considerable fractions of the neutron energy to the target nucleus. A neutron colliding head-on with a proton will give practically all its kinetic energy to the proton. As the mass of the target nucleus increases, the transfer of energy decreases, in accordance with the laws of conservation of energy and momentum. The loss of energy by mechanical impact is utilized in slowing down fast neutrons, a process known as *moderation.* Slow neutrons are most useful, for example, in the production of radioelements from stable elements by neutron capture. A good moderator should have low mass and a small capture cross section. The rate r of capture of neutrons from a neutron flux F (neutrons cm^{-2} sec^{-1}) incident on a layer of matter having N nuclei per square centimeter is given by

$$r = F\sigma N$$

where σ is the complete probability of capture. Replacing r by dN/dt and writing the flux as nv, where n is the number and v is the velocity of the neutrons, we have

$$dN/dt = nv\sigma N$$

which integrated gives

$$N = N_0 e^{-nv\sigma t}$$

where N is the number of unchanged nuclei in the target area at time t and N_0 is the number at time $t = 0$. The cross section σ is so named because it has the dimensions of an area. The unit for the cross section is the *barn*, equal to 10^{24} cm^2. When, as is often the case, σ is proportional to $1/v$, the advantage of slow neutrons in capture interactions becomes apparent. When the value of σ departs sharply from that predicted by the $1/v$ law, it usually increases over a narrow range of energies, and we have what is called a *resonance.* Slow neutron cross sections are customarily quoted for thermal neutrons at 20°C, corresponding to a value of v of 2200 m/sec. Under these conditions, representative thermal neutron capture cross sections are: boron, 759 barns; cobalt, 38 barns; cadmium, 2450 barns; gadolinium, 46,000 barns; gold, 99.8 barns; helium, 0; lead, 0.170 barn; and oxygen $<$0.0002 barn.

Additional interactions of neutrons with nuclei include the release of charged particles by neutron-induced nuclear disintegration. Commonly known reactions are *n-p*, *n-d*, and *n*-α. In these cases, the incident neutrons may contribute part of their kinetic energy to the target nucleus to effect the disintegration. Hence, more than mere neutron capture is involved. Then, there is usually a lower threshold for the neutron energy below which the reaction fails to occur. Another important reaction involving neutrons is fission, which may occur under different conditions for either slow or fast neutrons with appropriate fissionable material.

Sources of Neutrons. Any nuclear reaction in which neutrons are released might serve as a source of neutrons. In the initial experiments with neutrons, an α-*n* reaction was used. Because of the charge on the α-particle, it must have a high kinetic energy to penetrate a nucleus. Thus, polonium α-particles could release neutrons from beryllium. Such a natural source produces relatively few neutrons. The yield of neutrons from charged particle reactions can be increased manyfold by the use of particle accelerators. Here large numbers of charged particles of high energy can be used in the bombardment of the target to release numerous neutrons. Frequently deuterons or protons are used for the bombardment. A far more prolific source is the nuclear reactor. Fission of uranium is usually the source of the neutrons in this case. A nuclear reactor as usually constructed generates neutrons of different energies in various parts of its structure. Neutrons of suitable energy for a given experiment may be brought outside the reactor through channels into appropriate sections of the reactor. See also **Nuclear Power.**

Traditionally, the neutron is regarded as a particle which is a component of nuclei and which exists only briefly in the free state. For many purposes, this view is sufficient. However, it became obvious some years ago from various experiments, for example, in very high energy accelerators, that the neutron must have a complex structure. This view was reinforced by the nature of of the decay of the neutron. A ςparticle is ejected from the neutron on decay, but it is quite certain that the electron did not exist within the neutron prior to the decay. Rearrangements of an internal structure of the neutron must provide the energy for the formation and ejection of the ςparticle. An early theory would have the neutron consist of a proton and a π^- meson bound together so that they oscillate between a completely bound state and a more loosely bound state. This concept might explain the feeble interaction which has also been observed between electrons and neutrons at very short range.

Fermi Age Model. This is a model for the study of the slowing down of neutrons by elastic collisions. It is assumed that the slowing down takes place by a very large number of very small energy changes. Phenomena due to the finite size of the individual losses are ignored. In this model, the word age is somewhat of a misnomer, since its units are those of area rather than time. The name arises because the variable τ, the Fermi age, appears in the Fermi age equation in the same way that time appears in the standard heat-diffusion equation. The equation for the Fermi age in a unit volume of nuclear reactor is $\tau = D\phi/q$, in which D is the diffusion coefficient for fast neutrons, $\phi = nv$ is the fast neutron fluence, and q is the number of neutrons thermalized per second per cubic centimeter. For this purpose fast neutrons are inclusively all neutrons with energies between those acquired at fission and that energy at which they are thermalized.

Ultracold Neutrons. As pointed out by King (Massachusetts Institute of Technology), there is probably as much to be learned between the lowest energy yet reached and zero energy as there is between the highest energy attained and infinite energy. Ultracold neutrons may provide an avenue to very-low-energy research. It should be recalled that a neutron with an energy of 10^{-7} electron volt is at the low end of the energy scale. A neutron has the energy that would be imparted to an electron by a potential difference of one ten-millionth of a volt (0.1 microvolt). As pointed out by Golub et al. in 1979, this is the amount of energy of a particle in a gas whose temperature is one millidegree K. Unlike high-energy particles, ultracold neutrons move at a rate measured in a few meters per second. Golub et al. have proposed that inasmuch as ultracold neutrons cannot penetrate a solid surface, they can be confined in a metal bottle and by storing over long periods, it may be possible to measure the fundamental properties of the neutron.

Outside the nucleus, the neutron is an unstable particle. Free neutrons are rare in nature. By the process of beta decay, the neutron breaks down into a proton, an electron, and a neutrino (a massive particle). For probing atomic and molecular structures, thermal neutrons have traditionally been used. As explained by Golub et al., ultracold neutrons can be employed in a similar way, but their low energy and long wavelength adapt them to the examination of materials on a somewhat larger scale. At the Technical University (Munich), Steyeri and associates have developed a neutron spectrometer. In conventional neutron spectrometers, particles are analyzed by a magnet that bends their trajectories. In an ultracold-neutron spectrometer, the earth's gravitational field is used. In the device, the neutrons enter the spectrometer in a horizontal movement and are accelerated as they fall a fixed distance to a specimen. Those neutrons that rebound from a target are collected by an exit slit of the instrument. An exchange of energy with the specimen is reflected by the maximum height to which the neutrons rebound.

During the past decade, since their detection, much research has been directed toward methods of extracting, storing, and manipulating ultracold neutrons. It now appears that the next period will be one of investigating the neutron per se and possibly of using this new knowledge for the study of other systems of particles.

See **Particles (Subatomic)** for more recent views. Also see **Proton.**

Glossary of Neutron Terms

Neutrons are designated according to their energies, including the following:

Thermal neutrons, or neutrons in thermal equilibrium with the substance in which they exist; most commonly, neutrons of kinetic energy about 0.025 eV, which is about $\frac{2}{3}$ of the mean kinetic energy of a molecule at 15°C.

Epithermal neutrons, or neutrons having energies just above those of thermal neutrons; the epithermal neutrons energy range is between a few hundredths eV and about 100 eV.

Slow neutrons (a less definite classification), which may mean either neutrons having energies up to about 100 eV, or thermal neutrons.

Intermediate neutrons, which are neutrons having energies in a range that extends roughly from 100 to 100,000 eV. This range is above that of epithermal neutrons and below that of fast neutrons.

Fast neutrons, which are neutrons with energies exceeding 10^5 eV, although sometimes a lower limit is given.

Resonance neutrons may be either of the following: (1) for a specified nuclide or element, neutrons that have energies in the region where the cross section of the nuclide or element is particularly large because of the occurrence of a resonance. For example, cadmium resonance neutrons have energies between 0.05 and about 0.3 eV. (2) Neutrons having kinetic energies in the region of values for which prominent resonances are encountered in many nuclides; loosely, epithermal neutrons.

Prompt neutrons are those neutrons released coincident with a fission process.

Delayed neutrons are those neutrons released subsequently in a fission process, or, more generally, neutrons emitted by excited nuclei formed in any radioactive process (beta disintegration, in all cases so far known). The neutron emission itself is prompt, so that the observed half life is that of the preceding beta emitter. The situation is similar to that involving gamma-ray emission, which is a competing process. Delayed neutron emission is possible only if the excitation energy of the product nucleus exceeds the neutron binding energy for that nucleus. The chemistry of the delayed neutron emitter is that of the beta activity; thus ^{87}Br, ^{137}I, and ^{17}N are delayed neutron precursors, although the neutron emission actually takes place from excited nuclei of the products ^{87}Kr, ^{137}Xe, and ^{17}O.

Neutron cycle is the average life history of a neutron in a nuclear reactor. The gain in the number of neutrons in a reactor during any individual neutron cycle is given by $n(k - 1)$, where n is the number of neutrons in the reactor of the beginning of the cycle and k is the multiplication factor.

Neutron excess is the difference between the number of neutrons and the number of protons in an atomic nucleus. This is found by subtracting the atomic number of that nuclide from the neutron number; or by subtracting twice the atomic number from the mass number.

Neutron flux density is the number of neutrons that enter a sphere of unit cross-sectional area per unit of time. This quantity is sometimes defined in terms of a unidirectional beam of neutrons incident perpendicularly upon a unit area, but this definition is less general. It is also sometimes called neutron current density.

Neutron number is the number of neutrons in a nucleus. Its symbol is N. The neutron number for a given nuclide is equal to the difference between the mass number and the atomic number for that nuclide.

Additional Reading

Breuker, H., et al.: "Tracking and Imaging Elementary Particles," *Sci. Amer.*, 58 (August 1991).

Gutbrod, H., and H. Stöcker: "The Nuclear Equation of State," *Sci. Amer.*, 58 (November 1991).

Harding, A. K.: "Physics in Strong Magnetic Fields Near Neutron Stars," *Science*, 1033 (March 1, 1991).

Ruthen, R.: "Out of Its Field: How the Neutron Responds to A Field That Exerts No Force," *Sci. Amer.*, 26 (October 1989).

Traweak, S.: "Beamtimes and Lifetimes. The World of High Energy Physics," Harvard University Press, Cambridge, Massachusetts, 1988.

NEUTRON ACTIVATION ANALYSIS. This is a method of elemental analysis based upon the quantitative detection of radioactive species produced in samples via nuclear reactions resulting from neutron bombardment of samples. The neutron-induced reactions are of two main types: (1) those induced by very slow (thermal) neutrons, having energies of about 0.025 eV; and (2) those induced by fast neutrons having energies in the range of MeV. The method is used in two different forms. The purely instrumental form is fast and nondestructive and is based upon the quantitative detection of induced gamma-ray emitters by means of multichannel gamma-ray spectrometry. The amount of the element present is usually computed from the photopeak (total absorption peak) height or area of its gamma ray, or one of its principal gamma rays, compared with that of the standard. Where interferences from other induced activities are serious and cannot be removed by decay, spectrum subtraction, or computer solution, one must turn to the radiochemical separation method. Here the activated sample is put into solution and equilibrated chemically with measured amounts (typically 10 milligrams) of added carrier of each of the elements of interest, before chemical separations are carried out. The element to be detected needs then to be recovered in chemically and radiochemically pure form, but it need not be quantitatively recovered, since the carrier recovery is measured and the counting data are then normalized to 100% recovery. This form is slower, but it applies to pure beta emitters, as well as to gamma emitters, and it does eliminate interfering activities.

NEUTRON INTERFEROMETER. See **Gravitation.**

NEUTRON STARS. One of the end points in stellar evolution, in which a star of 1.4 to approximately 3 times the mass of the sun is supported by internal pressure generated by compressed "degenerate" neutrons.

Normal stars result from a balance of forces: the force of gravity pulling inward and a force generated by the energy of the nuclear reactions inside pushing outward. When the star exhausts its core hydrogen, it leaves the "main sequence" of normal stars and contracts under the force of gravity. The increased fusion that follows provides energy for the outer layers to swell, and the star becomes a "red giant." Later, after the helium is exhausted, stars still contain more than about 4 solar masses expand again to become "red supergiants."

When the supergiants have built all their nuclear fuel into iron, they explode as *supernovae* (q.v.). Those that have more than about 1.44 solar masses left after this stage are not stopped by the force of electron degeneracy that supports *white dwarfs* (q.v.). 1.44 solar masses is a theoretically derived value known as the "Chandrasekhar limit," after S. Chandrasekhar of the University of Chicago. Those stars that have more than 1.44 but less than about 3 solar masses (with the value of this upper limit being less accurately known than the value of the lower limit) will contract until the neutrons in the star resist being pushed together any further. This generates a pressure known as "neutron degeneracy pressure." The resulting object is a *neutron star*.

This situation was predicted by J. Robert Oppenheimer and colleagues in the 1930s. It comes from the application of quantum mechanical rules, including the Pauli exclusion principle and the Heisenberg uncertainty principle, to the neutrons. A neutron star contains approximately 2 solar masses in a volume only 20 km or so across, giving a density of billions of tons per cubic centimeter.

Neutron stars may have solid, crystalline crusts about 100 m thick. The neutron stars' atmospheres may be only a few centimeters thick. The concentration of a normal stellar magnetic field by gravitational collapse would correspond to a neutron star magnetic field of billions of gauss.

Though no neutron stars were known until 1968, neutron stars are now detected in various ways. Most of the known neutron stars are *pulsars* (q.v.), of which over 400 are now known. In these cases, we detect a beam of radio radiation, possibly emitted along the magnetic axis of the star, as it sweeps across space. Periods of rotation of the neutron stars in pulsars range from 1.4 ms to about 3 s. A very few of these pulsars have been detected in the optical or x-ray part of the spectrum. The best-studied example is the pulsar in the Crab Nebula, the remnant of the supernova explosion of 1054. This object is detected across the entire spectrum. In addition to its radio pulsations, its pulsations have been recorded in the x-ray region with the Einstein Observatory spacecraft and in the optical region with telescopes on earth. Another unusual pulsar, in a binary system, has been used as a test of the general theory of relativity. (See **Cosmology.**)

Other neutron stars appear in binary x-ray sources. Hundreds were mapped by NASA's High Energy Astronomy Observatory 1 in the late 1970s, and many were then studied individually by the Einstein Observatory, High Energy Astronomy Observatory 2. Perhaps the best studied had been discovered by an early x-ray spacecraft, UHURU. This source, Hercules X-1, shows rapid pulsations every 1.24 seconds, but the pulses turn on and off with a 1.7-day period. This longer period is presumably caused by eclipses by the primary object in the system, the variable star HZ Herculis, while the shorter period is presumably caused by rotation of the neutron star. A third period of about 35 days also exists in Hercules X-1.

The most unusual astronomical object now known is SS433, whose spectral lines vary with an 164-day period over a Doppler range corresponding to velocities of about 25% of the speed of light. SS433 is in our galaxy, and such a velocity in our galaxy is unprecedented. A 13-day periodicity has also been discovered in SS433, indicating that it is a binary system of that period. In leading models, an accretion disk of material has formed around the neutron star, and jets of gas are emitted above and below the plane of the disk, which is inclined to us. As the disk precesses, we alternately see each of the beams approaching and receding with the longer period of the system. It is the material in the beams that has the high velocities.

Discovery of neutron stars has spurred work in stellar evolution of binary systems. Each object in a binary system has a Roche lobe around it, within which the gravitational force of the object is strong enough to resist the tidal forces from outside objects. The Roche lobes define a figure-of-8 around the two members of a binary system. In some binaries, one or both objects swell to fill their Roche lobes, at which time material can flow from one object to the next. The exchange of mass can severely alter the evolution of the objects, and the relative masses of the objects can interchange. When one of the objects is a compact object like a neutron star, mass from the companion flowing through the neck of the figure-of-8 and falling on the neutron star can lead to a surface nuclear explosion that is a powerful emitter of x-rays.

Jay M. Pasachoff

NEWCASTLE DISEASE. Also sometimes called *fowl pest, pseudo fowl pest, Ranikhet disease*, and *avian pneumoencephalitis*, the disease is of viral origin. The disease was first noted in the Dutch East Indies. A series of outbreaks of the disease occurred in England near Newcastle-on-Tyne, from which the name of the disease was derived.

In his excellent reference, R. F. Gordon devotes several pages to the history and characteristics of Newcastle disease and summarizes the situation as of the mid- to late 1970s as follows: "By 1966, a fairly stable situation seemed to have developed in which fully virulent virus had become endemic in the tropics, milder disease in North America and western Europe and an intermediate form in Iran and Arab countries to the West. In 1968, an upsurge of the disease was reported in Iraq. Reports of disease, difficult to control then followed from Lebanon, Israel, Greece and in 1970 from England and Holland and in 1971 to other countries in western Europe. Workers have given the strain the designation Essex '70. In 1971 the United States reported cases of fully virulent Newcastle disease occurring along its southern border and designated such strains as velogenic viscotropic Newcastle disease (VVND). A comparison of the two types of isolate has been made. The Essex '70 type of virus has been observed to be highly pneumotropic and to spread rapidly while the VVND type of isolate seems on average, more lethal, more likely to give rise to visceral lesions and to spread more erratically. Although no distinction can accurately be made by laboratory tests, there is considerable epidemiological evidence to suggest their mode of spread is significantly different. By 1974, the world use of vaccine had increased greatly and in countries such as the UK where slaughter on infected farms was practiced until 1962, there had been heavy expenditure on control. Currently losses from the disease are not high in countries where extensive vaccination is practiced although virulent forms of the virus are now much more widespread than ever before."

NEWT (*Amphibia, Urodela*). Air-breathing salamanders which are at least partially aquatic in habits. The many species are found chiefly in the northern hemisphere.

NEWTON-COTES FORMULA. A method of numerical intergration. Assume that the integral

$$\int_a^b f(x)\, dx$$

may be approximated by

$$\int_a^b \phi(x)\, dx = A_0 y_0 + A_1 y_1 + \cdots + A_n y_n$$

where the quantities A_i are independent of y_i. By proper choice of these quantities, which may be found by the method of undetermined coefficients, the numerical result may be made very close to the true value of the integral. Special cases of the formula are the trapezoidal rule, the Simpson rule, and the Weddle rule.

NEWTONIAN FLUID. See **Colloid System; Fluid.**

NEWTONIAN LIQUIDS. See **Viscosity.**

NEWTONIAN TELESCOPE. A reflecting-type telescope having a 45° mirror located just inside the focus, so that the primary image is observed through a hole in the side of the tube. See also **Telescope.**

NEWTON RINGS. An interference phenomenon, easily observed by laying a slightly convex lens upon a flat glass plate. When the lens and plate are arranged so that monochromatic light is reflected at a suitable angle to the observer's eye, the point of contact is seen to be surrounded by a series of concentric, alternately bright and dark rings, which become closer together with increasing radius. The rings are due to the interference of light at the film of air between the glass surfaces, which film increases in thickness with increasing distance from the contact point. If the radius of curvature of the convex surface is R, and if, counting the central contact-spot as the zero ring, we number the rings in order, both bright and dark, from the center out, the radius of the Nth ring in monochromatic light of wavelength λ is approximately

$$a = \sqrt{\frac{NR\lambda}{2}}$$

With white light, the bright rings become colored spectra, the overlapping of which at larger values of N causes the system to become indistinct and disappear.

NEWTON'S FORMULA FOR INTERPOLATION. Let a difference table be given with numerical values of $y_0, y_1, y_2, \ldots$; equally spaced values of the argument, $x_0, x_1, x_2, \ldots$; $h = (x_n - x_0)/n$ and the finite differences $\Delta^n y_k$. Then a value of y for $x = x_k + hu$, not contained in the table, may be found by Newton's formula for forward interpolation

$$y = y_k + u\,\Delta y_k + \binom{u}{2}\,\Delta^2 y_k + \cdots + \binom{u}{2}\Delta^n y_k$$

As its name implies, this equation is used for calculation near the beginning of a difference table. Near the end of such a table, Newton's formula for backward interpolation is appropriate

$$y = y_k - v\,\Delta y_{k-1} + \binom{v}{2}\Delta^2 y_{k-2} - \cdots \binom{v}{k}\Delta^k y_0$$

where $x = x_k - hv$. These equations are also known as the Gregory-Newton formula.

See also **Interpolation.**

NEWTON'S LAWS OF DYNAMICS. The classical or Newtonian dynamics rests upon certain propositions first enunciated in systematic form by Sir Isaac Newton, which he set forth as three "Laws" of force and motion.

The first Law states, in effect, that bodies of matter do not alter their motions in any way except as the result of forces applied to them. A body at rest remains at rest, or if in motion it continues to move in the same direction with the same speed, unless a force is impressed upon it. It is quite conceivable that Newton's interpretation of "force" was the primitive concept which we all have, based on muscular effort, and that he regarded this statement as the expression of a natural law connecting force with motion. On the other hand, he may have recognized in this first Law, as we now do, an objective definition of force, namely, that which is capable of altering bodily motions in the face of an opposition called inertia whose nature is even now not fully understood.

The second Law is made up of two distinct parts: (1) When different forces are allowed to act upon free bodies, the rates at which the momentum changes are proportional to the forces applied. (2) The direction of the change in momentum caused by a force is that of the line of action of the force. Part 1 may now be regarded, from the more rigorous viewpoint of dimensional analysis, as a definition of the standard measure of force. Two forces are judged equal if they produce change of momentum at equal rates; one force is twice as great as another if it changes the momentum at twice the rate; etc. Moreover, two bodies have equal mass if equal forces produce change of momentum at equal rates; and one body has twice the mass of another if an equal force changes its momentum at half the rate. This is the inertial concept of mass mentioned above. Part 2 emphasizes the vector character of force, and points out that no single force, acting alone, can cause a change of motion in any direction save that of its own line of action. If the effect is apparently in some other direction, as when a string operates over a pulley, the force is always combined with one or more auxiliary forces, the resultant of all of them being in the direction of the observed change of motion.

The third Law asserts the equality of "action and reaction." In the case of forces acting on bodies at rest, the principle is easily illustrated. When a steel truss rests on a pier and presses downward upon it with a force of 100,000 units, the pier exerts an upward thrust or "reaction" against the truss, also of 100,000 units, and this thrust, tending to bend the truss upward, is a most important factor in computing the stresses in the truss members. The Law also applies to forces acting upon bodies free to yield and to receive acceleration; a fact not explicitly stated by Newton, and discussed more fully elsewhere as d'Alembert's principle.

These propositions constituted the unquestioned foundation of dynamics until about the beginning of the twentieth century. So far as practical operations with bodies of ordinary size are concerned, they still answer every purpose. It is only when we consider motions with velocities comparable to that of light, or attempt to analyze the mechanics of bodies of the atomic and electronic order of magnitude, that the Newtonian dynamics breaks down and must be replaced by a system founded upon the postulates of relativity and the concepts of the quantum theory.

NEW WORLD. A term used in some of the natural sciences, essentially to designate the Western Hemisphere.

NEYMAN-PEARSON THEORY. In statistical inference, the theory of tests of hypotheses developed by J. Neyman and E. S. Pearson. It is based on the delimitation of two types of error, the rejection of a true hypothesis and the acceptance of a false hypothesis. Errors of the first kind are controlled at assigned probability levels. Errors of the second kind cannot simultaneously be so controlled, but can be explored for different values of parameters alternative to the true one. Tests which have a smaller probability of the second kind of error are said to be more powerful and part of the theory is devoted to seeking for most powerful tests. The probability of rejection of a false value of a parameter, graphed as ordinate against the value of the parameter as abscissa, is the Power Function of the test.

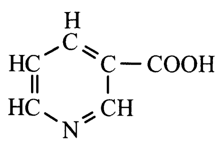
NIACIN. Sometimes referred to as nicotinic acid or nicotinamide and earlier called the P-P factor, antipellagra factor, anti-blacktongue factor, and vitamin B_4, niacin is available in several forms (niacin, niacinamide, niacinamide ascorbate, etc.) for use as a nutrient and dietary supplement. Niacin is frequently identified with the B complex vitamin grouping. Early in the research on niacin, a nutritional niacin deficiency was identified as the cause of pellagra in humans, blacktongue in dogs, and certain forms of dermatosis in humans. Niacin deficiency is also associated with perosis in chickens as well as poor feathering of the birds.

Varying in degree in relationship to the length and severity of diet deficiency of niacin, pellagra is clinically manifested by skin, nervous system, and mental conditions. The disease occurs most frequently among the economically deprived and particularly in areas where the diet may be high in maize (corn) intake. The disease was first described by Gaspar Casal in 1735 and was common in many areas, including Europe, Egypt, Central America, and the southern portion of the United States for many years. The largest outbreak occurred in the United States during the period 1905–1915 and resulted in a high mortality. The medical awareness and understanding of vitamins and dietary deficiencies, coupled with the availability of dietary supplements in staple foods, have resulted in a great lessening in the occurrence of pellagra. Where the disease is found, niacin is a specific for the treatment of acute pellagra. Those afflicted are accustomed to a diet low in protein and made up largely of carbohydrates. Predisposing causes are found idiosyncrasies, chronic alcoholism, and diseases which interfere with the assimiliation of a proper diet.

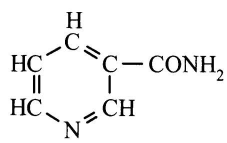
Huber first synthesized nicotinic acid in 1867. In 1914, Funk isolated nicotinic acid from rice polishings. Goldberger, in 1915, demonstrated that pellagra is a nutritional deficiency. In 1917, Chittenden and Underhill demonstrated that canine blacktongue is similar to pellagra. In 1935, Warburg and Christian showed that niacinamide is essential in hydrogen transport as diphosphopyridine nucleotide (DPN). In the following year, Euler et al. isolated DPN and determined its structure. In 1937, Elvhehjem et al. cured blacktongue by administration of niacinamide derived from liver. In the same year, Fouts et al. cured pellagra with niacinamide. In 1947, Handley and Bond established conversion of tryptophan to niacin by animal tissues.

In the physiological system, niacin and related substances maintain nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP). Niacin also acts as a hydrogen and electron transfer agent in carbohydrate metabolism; and furnishes coenzymes for dehydrogenase systems. A niacin coenzyme participates in lipid catabolism, oxidative deamination, and photosynthesis.

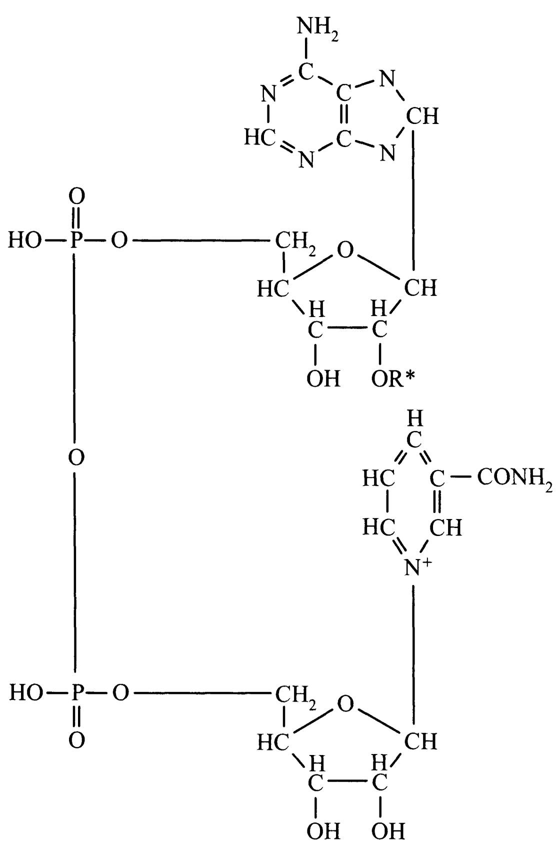
Nicotinic acid can be converted to nicotinamide in the animal body and, in this form, is found as a component of two oxidation-reduction coenzymes, NAD and NADP, as previously mentioned. Structurally, these are:



Nicotinic acid



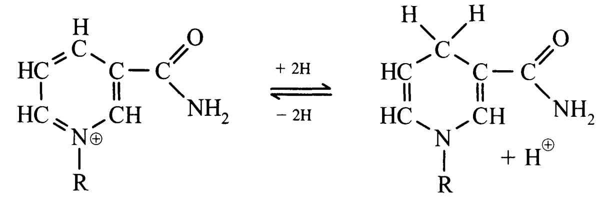
Nicotinamide



Nicotinamide adenine dinucleotide (NAD) R* = H

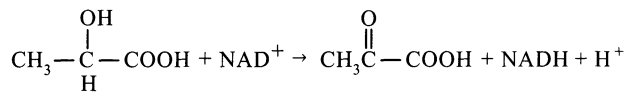
Nicotinamide adenine dinucleotide phosphate (NADP) R* = $-\overset{\displaystyle O}{\overset{\|}{P}}-OH$ with OH below P ($R^* = -P(=O)(OH)-OH$)

The nicotinamide portion of the coenzyme transfers hydrogens by alternating between an oxidized quaternary nitrogen and a reduced tertiary nitrogen as shown by:



NAD (oxidized) — NADH + H $^+$ (reduced)

Enzymes that contain NAD or NADP are usually called dehydrogenases. They participate in many biochemical reactions of lipid, carbohydrate, and protein metabolism. An example of an NAD-requiring system is lactic dehydrogenase which catalyzes the conversion of lactic acid to pyruvic acid. Numerous NAD-dependent enzyme systems are known.



Lactic acid — Pyruvic acid

Distribution and Sources. In plants, niacin production sites occur in leaves, germinating seeds, and shoots. In humans, niacin is not available from intestinal bacteria, but some conversion is made from tryptophan which occurs in tissues.

High niacin content (10–100 milligrams/100 grams) Chicken (white meat), groundnut (peanut), halibut, heart (calf), kidney (beef, pork), liver (beef, calf, chicken, pork, sheep), meat extracts, rabbit (white meat), swordfish, tuna, turkey (white meat), yeast

Medium niacin content (1–10 milligrams/100 grams) Almond (dry), asparagus, avocado, barley, bean (kidney, lima, snap, wax), beef, broccoli, cashew, cheeses (camembert, roquefort, Swiss), chestnut, chicken (dark meat), clam, date (dry), duck, fig (dry), fishes (except those listed under "High"), kale, lamb, lentil (dry), maize (corn), molasses, mushroom, oats, oyster, parsley, pea, potato, prune (dry), rice (brown), rye, soybean (dry), shrimp, walnut, wheat, wheat germ

Low niacin content (0.1–1.0 milligram/100 grams) Apple, apricot, banana, beet, beet greens, berries (black-, blue-, cran-, rasp-, straw-), Brussels sprouts, cabbage, carrot, cauliflower, celery, cherry, chicory, coconut, cucumber, currant, dandelion greens, eggs, eggplant, endive, fig, grape, kohlrabi, lettuce, lemon, melons, milk, onion, parsnip, peach, pear, pecan, pepper, pineapple, plum, pumpkin, radish, raisin (dry), rhubarb, spinach, sweet potato, tangerine, tomato, turnip, watercress

Precursors in the biosynthesis of niacin: In animals and bacteria, tryptophan; and in plants, glycerol and succinic acid. Intermediates in the synthesis include kynurenine, hydroxyanthranilic acid, and quinolinic acid. In animals, the niacin storage sites are liver, heart, and muscle. Niacin supplements are prepared commercially by: (1) Hydrolysis of 3-cyanopyridine; or (2) oxidation of nicotine, quinoline, or collidine.

Bioavailability of Niacin. Factors which cause a decrease in niacin availability include: (1) Cooking losses; (2) bound form in corn (maize), greens, and seeds is only partially available; (3) presence of oral antibiotics; (4) diseases which may cause decreased absorption; (5) decrease in tryptophan conversion as in a vitamin B_6 deficiency. Factors which increase availability include: (1) Alkali treatment of cereals; (2) storage in liver and possibly in muscle and kidney tissue; and (3) increased intestinal synthesis.

Antagonists of niacin include pyridine-3-sulfonic acid (in bacteria); 3-acetylpyridine, 6-aminonicotinamide, and 5-thiazole carboxamide. Synergists include vitamins B_1, B_2, B_6, B_{12}, and D, pantothenic acid, folic acid, and somatotrophin (growth hormone).

In humans, overdosage of niacin causes a limited toxicity (1 to 4 grams/kilogram) with individual variations in sensitivity.

NIACINAMIDE. See **Pyridine and Derivatives.**

NICHE. See **Ecology.**

NICKEL. Chemical element, symbol Ni, at. no. 28, at. wt. 58.69, periodic table group 10, mp 1453°C, bp 2732°C, density 8.9 g/cm^3 (solid, 20°C), 9.04 g/cm^3 (single crystal). Elemental nickel has a face-centered cubic crystal structure. Nickel is a silver-white metal, harder than iron, capable of taking a brilliant polish, malleable and ductile, magnetic below approximately 360°C. When compact, nickel is not oxidized on exposure to air at ordinary temperatures. The metal is soluble in HNO_3 (dilute), but becomes passive in concentrated HNO_3. The metal does not react with alkalis. Finely divided nickel dissolves 17 × its own volume of hydrogen at standard conditions. There are five naturally occurring stable isotopes, ^{58}Ni, ^{60}Ni through ^{62}Ni, and ^{64}Ni. Six radioactive isotopes have been identified, ^{56}Ni, ^{57}Ni, ^{59}Ni, ^{63}Ni, ^{65}Ni, and ^{66}Ni. ^{59}Ni has a half-life of 8×10^4 years, and ^{63}Ni has a half-life of 80 years. The half-lives of the remaining radioactive isotopes are relatively short, expressed in hours and days. The element ranks 21st among the elements in terms of abundance in the earth's crust, the estimated average content of igneous rocks being about 0.02%. In terms of cosmic abundance, nickel ranks 28th among the elements. Nickel ranks 40th in terms of concentration in seawater, the estimated content being about 2.5 tons of nickel per cubic mile (540 kilograms per cubic kilometer) of seawater. Awareness of nickel probably dates back to antiquity, but the element was not firmly identified until 1751 when Axel Fredric Cronstedt isolated the metal from the sulfide ore NiAsS.

First ionization potential 7.33 eV, second, 18.13 eV. Oxidation potentials $Ni \rightarrow Ni^{2+} + 2e^-$, 0.230 V; $Ni^{2+} + 2H_2O \rightarrow NiO_2 + 4H^+ + 2e^-$, −1.75 V; $Ni + 2OH^- \rightarrow Ni(OH)_2 + 2e^-$, 0.66 V; $Ni(OH)_2 + 2OH^- \rightarrow NiO_2 + 2H_2O + 2e^-$, −0.49 V.

Other physical properties of nickel are given under **Chemical Elements.**

In the early 1800s, the principal sources of nickel were in Germany and Scandinavia. Very large deposits of lateritic (oxide or silicate) nickel ore were discovered in New Caledonia in 1865. The sulfide ore deposits were discovered in Sudbury, Ontario in 1883 and, since 1905, have been the major source of the element. The most common ore is pentlandite, $(FeNi)_9S_8$, which contains about 34% nickel. Pentlandite usually occurs with pyrrhotite, an iron-sulfide ore, and chalcopyrite, $CuFeS_2$. See also **Chalcopyrite; Pentlandite; Pyrrhotite.** The greatest known reserves of nickel are in Canada and Russia, although significant reserves also occur in Australia, Finland, the Republic of South Africa, and Zimbabwe.

Principal producers and/or exporters of nickel include, in diminishing order, Canada, Russia, the United Kingdom, Norway, and Indonesia. Main consumers are the United States, Japan, the United Kingdom, Norway, Germany, Canada, and France.

After beneficiation of the raw ore to form a sulfide concentrate, the latter is roasted to achieve partial oxidation of iron and partial removal of sulfur. The roasted material then is smelted with a flux to eliminate the rock content. At this point, part of the iron goes into the slag. The remaining material is a copper-bearing nickel-iron matte, made up mainly of the sulfides of these metals. The matte is then treated in a Bessemer converter to achieve further removal of iron and sulfur. After controlled cooling, which assists separation, the Bessemer product is finely ground and subjected to magnetic separation and differential flotation. The separated product is an impure nickel sulfide. The sulfide then is sintered to nickel oxide. This product may be marketed for some applications, but the majority of the oxide is cast into anodes for refining into nickel metal by one of two major processes.

In (1) the electrolytic process, a nickel of 99.9% purity is produced, along with slimes which may contain gold, silver, platinum, palladium, rhodium, iridium, ruthenium, and cobalt, which are subject to further refining and recovery. In (2) the Mond process, the nickel oxide is combined with carbon monoxide to form nickel carbonyl gas, $Ni(CO)_4$. The impurities, including cobalt, are left as a solid residue. Upon further heating of the gas to about 180°C, the nickel carbonyl is decomposed, the freed nickel condensing on nickel shot and the carbon monoxide recycled. The Mond process also makes a nickel of 99.9% purity.

Uses. The three main commercial forms of primary nickel are: (1) electrolytic sheets, (2) pellets resulting from the decomposition of nickel carbonyl, and (3) ferronickel. Traditionally, pellets are favored in Europe, whereas electrolytic nickel is favored in North America. Additional forms of commercial nickel are powder, ingots, shot, and briquettes. Ferronickel, containing 24–48% nickel with the remainder iron, is used mainly in the production of stainless steel. More than half of the nickel produced is used in stainless steels and high-nickel alloys. Additional uses include nickel plating, iron and steel castings, coinage, and copper and brass products.

The main consumer of nickel is austenitic stainless steel which contains from 3.5 to 22% nickel and 16 to 26% chromium. In these steels, nickel stabilizes the austenite and enhances the ductility of the steel. Nickel, along with chromium, contributes to corrosion resistance. Up to amounts of about 9%, nickel adds strength, hardness, and toughness to many alloy steels. Alloys in the 9% nickel range remain stable at low temperatures and are capable of handling liquefied gases. The lower-nickel steels (0.5 to 0.7%) are ductile, strong, and tough, and find use for many automobile parts, in power machinery, and construction equipment. There are hundreds of nickel-containing alloys, running the gamut from hardenable silver alloy (0.02% Ni) up to malleable nickel (99% Ni).

Wrought Nickel and High-Nickel Alloys. Some of the major nickel alloys, along with wrought nickel, are described in the accompanying table.

Commercially pure wrought nickel in the form of sheets, wire, and tubing has many uses because of its corrosion resistance. These uses include utensils, food-processing equipment, marine hardware, coinage, and chemical equipment. Electroplated nickel also is used as a protective coating on steel. *Nimonic* alloys, not shown on the table, are based on an 80% Ni, 20% Cr composition. They are high-strength, heat-resistance metals that are age-hardened to increase strength at elevated temperatures—with a useful range of 700–825°C. *Monel* metal (several types) is a high-strength corrosion-resistant alloy available in many wrought and cast forms for use in processing equipment, marine construction, and household appliances. *K Monel* can be heat treated by precipitation hardening to about 2 × the strength of annealed *Monel. Hastelloy*-type alloys are well known for their excellent resistance to HCl, H_2SO_4, and other acids. The *Incoloy*-type alloys (35% Ni approximately) are heat-resistant alloys used mainly as castings for furnace parts. The lower-nickel/higher-chromium alloys generally are classified as stainless steels. See also **Iron Metals, Alloys, and Steels.**

Although not of high-tonnage production, several nickel metals serve important uses, such as:

Permalloy, 78.5% Ni, 21.5% Fe; *Hipernik*, 50% Ni, 50% Fe; and *Perminvar*, 45% Ni, 30% Fe, 25% Co—are representative of a group of high-nickel magnetic alloys.

Constantan, 45% Ni, 55% Cu, has high electrical resistivity and a very low temperature coefficient of resistivity. It is extensively used with copper as a thermocouple element.

Nichrome, 80% Ni, 20% Cr (several types with variations of these percentages and additions of other elements, such as silicon in small amounts), is used as resistance wire for heating elements.

Calorite, 65% Ni, 8% Mn, 12% Cr, 15% Fe, also is used in electric heating elements.

Alumel, 94% Ni, 2.5% Mn, 0.5% Fe plus small amounts of other elements, is used in thermocouples

Chromel, 35–60% Ni, 16–19% Cr, generally with the balance Fe, also is used as resistance wire and for thermocouples.

Invar, 36% Ni, 64% Fe, has a very low temperature coefficient of expansion and is used for measuring tapes, instruments, and bimetallic thermostats.

Elinvar, 34% Ni, 57% Fe, 4% Cr, 2% W, has a very low temperature coefficient of elasticity which makes it useful for springs in watches and precision instruments.

There are hundreds of special nickel-bearing alloys of proprietary formulations and tradenames.

WROUGHT NICKEL AND REPRESENTATIVE NICKEL ALLOYS

	Melting Range °C	Poisson's Ratio
Wrought nickel 99% Ni, 0.25% Cu, 0.15%C	1,435–1,445	0.31
Duranickel 301 93.9% Ni, 0.05% Cu, 0.15% C, 0.15% Fe, 0.5% Ti, 4.5% Al	1,400–1,440	0.31
Monel 400 66.0% Ni, 31.5% Cu, 0.12% C, 1.35% Fe	1,300–1,350	0.32
Hastelloy B 63.5% Ni, 0.05% C, 5.0% Fe, 2.5% Co, 1.0% Cr, 28.0% Mo, 0.3% V	1,320–1,460	—
Hastelloy F 45.5% Ni, 0.05% C, 20.5% Fe, 2.5% Co, 22.0% Cr, 6.5% Mo, 1% W, 2% (Nb + Ta)	1,290–1,295	0.305
Inconel 600 72% Ni, 0.5% Cu, 0.15% C, 8.0% Fe, 15.5% Cr	1,370–1,425	0.29
Incoloy 800 32.5% Ni, 0.75% Cu, 0.10% C, 45.6% Fe, 21.0% Cr	1,355–1,390	0.30
Illium G 56.0% Ni, 6.5% Cu, 22.5% Cr, 6.5% Mo	1,255–1,340	0.29

NOTE: Recently introduced new or improved nickel alloys include:

Inconel alloy 625—Low-cycle fatigue resistance has been increased from 70–80,000 to 110–120,000 psi at 10 cycles. This has been achieved through grain size control and improved product cleanliness. Major applications are bellows and expansion joints.

Inconel alloy 725—An age-hardenable alloy for deep sour gas well service, combining height strength with the attributes of Inconel alloy 625, such as pitting resistance and stress corrosion cracking resistance to salt, hydrogen sulfide, and sulfur at temperatures up to about 230°C (450°F) and to sulfide stress corrosion cracking.

Inconel alloy 622—Modified composition and special thermal mechanical processing give this alloy superior thermal stability and resistance to intergranular attack and localized corrosion. The alloy is particularly suited to acidified halide environments, especially those containing oxidizing acids.

Inconel alloy 925—An age-hardenable nickel-iron-chromium alloy providing high strength up to 540°C (1000°F). Developed for use in gas production applications, such as tubular products, tool joints, and equipment for surface and downhole hardware in gas industry.

Inco alloy 25–6MO—Used for its corrosion resistance in many environments, this is an austenitic nickel-iron-chromium alloy with a substantial (6%) addition of molybdenum. Especially useful for resisting pitting and crevice corrosion in media containing chlorides or other halides. Applications include equipment for handling sulfuric and phosphoric acids, offshore platforms and other marine equipment, and for bleaching circuits in pulp and paper plants.

Alloy with Memory. In seeking a way to reduce the brittleness of titanium, U.S. Navy researchers serendipitously discovered a nickel-titanium alloy having an amazing memory. Previously cooled clamps made of the alloy (*nitinol*) are flexible and can be placed easily in position. When warmed to a given temperature, the alloy hardware then exerts tremendous pressure. Use of conventional clamps for holding bundles of wires or cables in a ship or aircraft structure requires special tools. For this and other applications in industry and medicine, nitinol has been in demand. The alloy, however, is not easy to produce because only minor variations in composition can affect the "snap back" temperature by several degrees of temperature.

Nickel Powders. The use of nickel in powdered form has increased markedly during the last few years. As shown by Fig. 1, nickel powders are available in several types and are used in a variety of products.

Nickel in Nanometer Materials. Coating a metal with an ultrathin layer of another metal creates properties not found separately in either of the materials. Considerable recent research has been directed toward improving the mechanical properties of bimetallic laminates, sometimes called *composition modulated films,* which have interlayers only a few nanometers thick. Attractive properties also have been found for similar systems, called *nanometer materials*. Nickel has been used in combination with copper, ruthenium, and other metals for producing these new materials.

Production of High-Performance Nickel Alloys. In the production of high-performance alloys, the critical first step of alloying requires sophisticated equipment, stringent controls, and expertise. Several production methods are used.

Air melting in electric-arc or induction furnaces is used for many alloys, sometimes for final alloying, with further refining by argon-oxygen decarburization. Melting in air can result in impurities in some alloys, a problem eliminated by vacuum induction melting, used to produce ingots for direct rolling or for remelting. Remelting is accomplished by two methods, both with precise, computerized control. *Electroslag* remelting uses electrical resistance heating to remelt an ingot (electrode) under molten slag containing fluxes that remove impurities. *Vacuum arc* melting refines the structure of cast electrodes in a contaminant-free chamber. Remelting yields alloys of the highest level of refinement.

Nickel Chemistry and Compounds

With its $3d^84s^2$ electron configuration, nickel forms Ni^{2+} ions. Having a nearly complete $3d$ subshell, nickel does not yield a $3d$ electron as readily as iron and cobalt, and trivalent and tetravalent forms are known only in the hydrated oxides, Ni_2O_3 and NiO_2, and a few complexes.

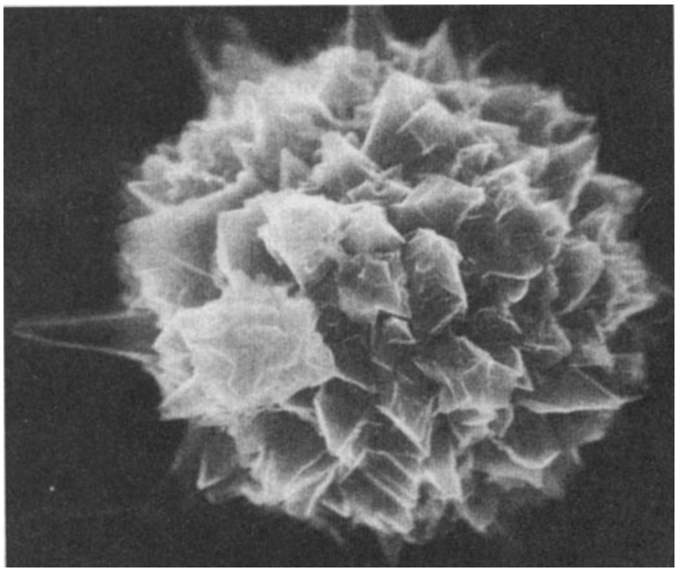
Nickel(II) oxide, NiO, produced by heating the carbonate, is thermally stable. Higher oxides of nickel, including Ni_2O_3 and NiO_2, are known only as hydrates, being prepared by vigorous oxidation of NiO in alkaline solution.

Nickel(II) sulfide, precipitated from Ni^{2+} solutions by ammonium sulfide, may show quite a little departure from stoichiometric composition. Like iron(II) and cobalt(II) FeS and CoS, it has in crystal form an electrical conductivity and other properties similar to a metal or alloy. There is no conclusive evidence that Ni_2S_3 can be prepared, but NiS_2 is known and believed to be, like FeS_2, a compound of Ni^{2+} and the S_2^{2-} ion.

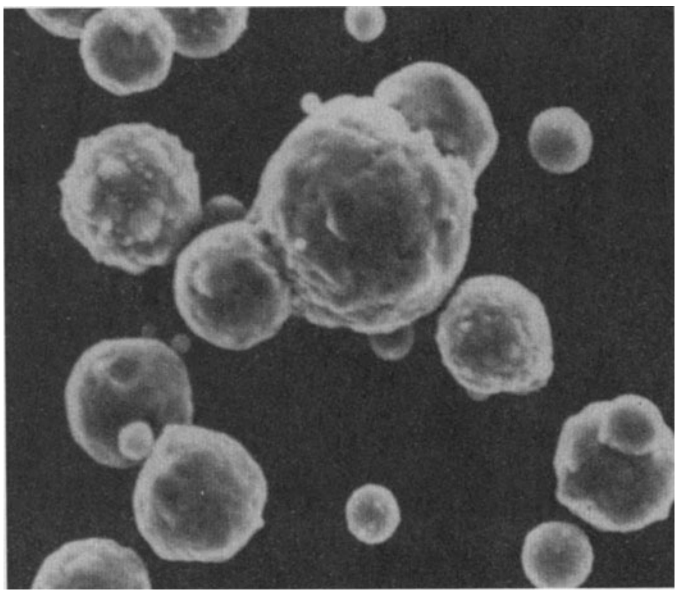
All four dihalides of nickel with the common halogens are known: NiF_2, formed by reaction of hydrofluoric acid or nickel(II) chloride or by thermal decomposition of $[Ni(NH_3)_6][BF_4]_2$, is greenish yellow, while the other three dihalides, formed directly from the elements, are green for the chloride, yellow for the bromide, and black for the iodide. In general, anhydrous Ni^{2+} salts are yellow and the ion $Ni(H_2O)_6^{2+}$ in aqueous solution is green.

Other elements with which nickel forms binary compounds, especially at higher temperature, are boron, carbon, nitrogen, silicon, and phosphorus. Like NiO, these compounds may depart slightly or even considerably from daltonide composition, frequently being interstitial compounds, and with higher elements of transition groups 5 and 6, merging into the interstitial compound-solid solution picture which nickel exhibits with the other transition metals.

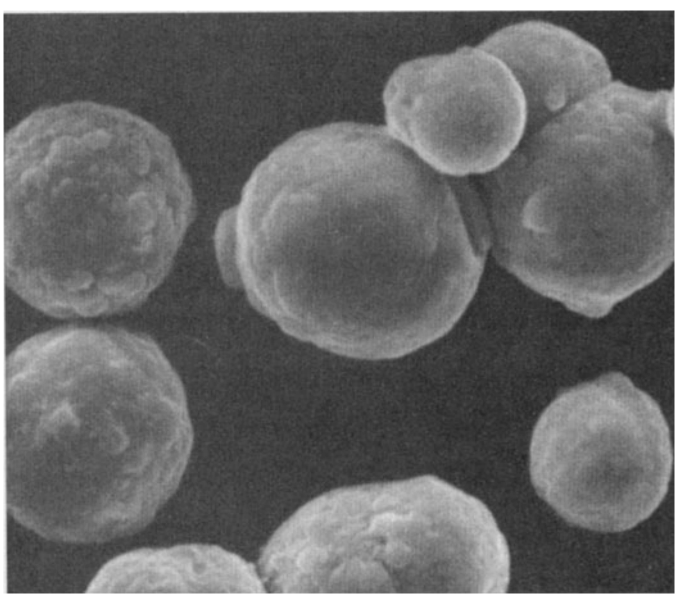
Divalent nickel forms two main types of complexes. The first consists of complexes of the spin-free ("ionic" or outer orbital) octahedral type (see **Ligand** for their discussion) in which the ligands are principally H_2O, NH_3, and various amines such as ethylenediamine and its derivates, e.g., $Ni(H_2O)_6^{2+}$, $Ni(NH_3)_6^{2+}$, $Ni(en)_6^{2+}$. These complexes usually have colors toward the high-frequency side of the spectrum, i.e., violet, blue, and green. The other class consists of tetracovalent square complexes with ligands such as CN^-, the dioximes and their deriva-



(a)



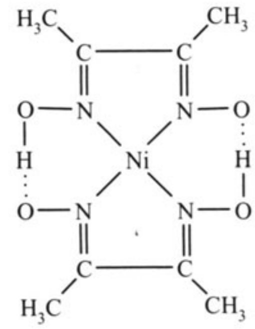
(b)



(c)

Fig. 1. Types of available nickel powders. (a) With a surface area of 0.4 m^2/g, this is a spiked nickel powder of single particles 3–7 microns in diameter, with a bulk density of 2 g/cc. The powder is used in both metal and chemical systems for powder metal parts, getters, magnets, electronic strip, flake, and organo-nickel compounds and nickel salts and soaps. (b) High-density nickel powder consisting of 8–12 micron semi-smooth particles, offering mixability with both metallic and nonmetallic powder systems. Applications include welding rods, nickel aluminide, nickel-columbium additives, abradable seals, powder metal parts, carbide binders, and conductive plastics. (c) Spherically shaped, high- purity nickel powder, with a surface area of 0.15 m^2/g and a Fisher particle size of 8–9 microns. Applications include friction materials, plasma spraying, metal injection molding, welding electrodes, magnets, cemented carbides, and powder metal nickel steels. (*Source: INCO Specialty Powder Products, Saddle Brook, New Jersey.*)

tives, and other chelates, which usually have colors on the low frequency side of the spectrum, i.e., red, orange, and yellow. The structure of the nickel-dimethylglyoxime complex is



This compound is of interest not only in analysis, but because by limited oxidation with the halogens it yields a unipositive ion containing trivalent nickel and also because the hydrogen bonds formed to the oxygen atoms are among the shortest known. Similarly, the tetracyanide complex of nickel, $Ni(CN)_4^{2-}$, may be reduced by sodium amalgam to give an ion of composition $Ni(CN)_4^{3-}$, or $(NC)_3Ni—Ni(CN)_3^{4-}$ containing Ni(I). This latter ion forms a potassium salt of nickel(I) of the formula $K_4Ni_2(CN)_6$ which is reduced in liquid NH_3 by metallic potassium to give the compound $K_4Ni(CN)_4$ in which the nickel has an effective valence of zero. Of course, this zero valence also exists in the carbonyls of nickel (and other elements) which, however, are covalent. $Ni(CO)_4$ is prepared by reaction of carbon monoxide with freshly reduced nickel, which occurs at ordinary temperatures and pressures. As with the carbonyls of other metals, the CO groups may be directly or indirectly, partially or completely, replaced by other groups. Derivatives of trivalent phosphorus form many such compounds of general formula $Ni(CO) \times {}_{4-x}(PR_3)_x$, where R may be one or more of such groups as F, Cl, Br, I, alkyl, aryl, alkoxy, aryloxy, etc.

See also **Nickel (In Biological Systems).**

Additional Reading

Carter, G. F., and D. E. Paul: "Materials Science and Engineering," ASM International, Materials Park, Ohio, 1991.

Hanson, A.: "The Metal That Remembers," *Technology Review (MIT)*, 26 (May–June 1991).

Houston, J., and P. Feibelman: "Ultrathin Metal Coatings Yield Unique Properties," *Advanced Materials & Processes*, 31 (March 1991).

Lancaster, J. R., Jr., Editor: "The Bioinorganic Chemistry of Nickel," VCH Publishers, New York, 1988.

Sax, N. R., and R. J. Lewis, Sr.: "Dangerous Properties of Industrial Materials," 7th Edition, Van Nostrand Reinhold, New York, 1989.

Staff: "A Quick Reference Guide to Nickel and High-Nickel Alloys," *Advanced Materials and Processes*, 54 (October 1991).

Staff: "Metals Forecast," *Advanced Materials & Processes*, 17 (January 1991); 24 (January 1992); 18 (January 1993).

Staff: "ASM Handbook—Properties and Selection: Nonferrous Alloys and Pure Metals," ASM International, Materials Park, Ohio, 1991.

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

NICKEL BATTERY. See **Battery.**

NICKEL (In Biological Systems). Despite its many pharmacologic and in vitro actions, convincing evidence showing that nickel is an essential element for some animal species did not appear until the early 1960s. There has been considerable further and more convincing research during the 1970s and early 1980s. An excellent summary of the status as of the late 1970s is given in the Spears-Hatfield reference.

Like most trace elements, nickel can activate various enzymes in vitro, but no enzyme has been shown to require nickel, specifically, to be activated. However, urease has been shown to be a nickel metalloenzyme and has been found to contain 6 to 8 atoms of nickel per mole of enzyme (Fishbein et al., 1976). RNA (ribonucleic acid) preparations from diverse sources consistently contain nickel in concentrations many times higher than those found in native materials from which the RNA is isolated (Wacker-Vallee, 1959; Sunderman, 1965). Nickel may serve to stabilize the ordered structure of RNA (Fuwa et al., 1960). Nickel may have a role in maintaining ribosomal structure (Tal, 1968, 1969). These studies and other information have led to the suggestion that nickel may play a role in nucleic acid and/or protein metabolism.

Nickel also may act to stimulate or inhibit the release of various hormones (Nielsen, 1971, 1972; Dormer et al., 1973; Clay, 1975; Horak-Sunderman, 1975). Nickel has been found to inhibit insulin release from the pancreas (Dormer et al., 1973; Clay, 1975), and stimulates glucagon secretion (Horak-Sunderman, 1975).

Nickel as an essential element in ruminant nutrition has not been proved conclusive as of the early 1980s. However, with nonruminants, some evidence indicates that certain species fed low-nickel diets have a greater infant mortality rate and a general degradation of the reproductive process (Nielsen, 1975; Anke et al., 1973).

Zinc and nickel appear to behave similarly at certain sites in the biological system. Both elements are capable of activating certain enzymes; for example, arginase is an enzyme which can be activated by either element (Parisi-Vallee, 1969). Stimulation of enzyme activity is at a site at which trace element substitutions or interactions may occur. However, some sacrifice of activity usually results when normally occurring metal is replaced by a trace metal. Nucleic acids as well as the ribosomes are likely sites of interaction between nickel and zinc. Both metals are consistently found in high concentrations firmly bound to RNA. It has been suggested that they function in maintaining the structure of RNA, thus preventing conformational changes. Nickel appears to be as effective as zinc at equal concentrations in this respect. Nickel and zinc are also found in ribosomal ash and studies have indicated that both can contribute to ribosomal conformation. The white blood cell is another possible site at which nickel and zinc may interact. Leukocytes are high in zinc and total leukocyte counts as well as differential white cell counts change drastically during a zinc deficiency. The interrelationship between nickel and zinc has been studied in vitro primarily in swine and rats. Their relationship has been studied largely from a substitution standpoint. Nickel appears to substitute for zinc to a certain extent in both species.

Similarly, the relationship between nickel and copper has been under study. One of the major functions of copper is in hemoglobin formation. Hemoglobin and hematocrit values decline rapidly during a copper deficiency. Copper is currently believed to exert its effect on hemoglobin metabolism through ceruloplasmin. Early work also indicated that nickel might be involved in hematopoiesis. Investigators in 1974 found a decreased concentration of copper in the lung and spleen of rats receiving 5 parts per million of nickel in drinking water. High levels of dietary nickel in rats and mice have been reported to decrease the activity of cytochrome oxidase, a copper-containing enzyme.

As pointed out by Eskew, Welch, and Cary in 1983, in contrast with the situation in animals, for which four new essential trace elements were identified in recent years, no new generally essential micronutrient for higher plants has been found since the mid-1950s. When it was found that urease is a nickel-metalloenzyme, this suggested that Ni may play a role in higher plants. Nickel has evidenced a stimulation of growth when urea is the sole source of nitrogen, but has slight or no effect when other nitrogen enrichment sources are used. The aforementioned investigators claim that Ni is essential for nitrogen metabolism in soybeans [*Glycine max* (L.) Merr.], either when nitrogen is furnished as NO_3^- and NH_4^+ or when plants depend upon nitrogen fixation. In experiments, soybean plants deprived of Ni accumulated toxic concentrations of urea (2.5%) in necrotic lesions on their leaflet tips. This occurred regardless of whether the plants were furnished with inorganic N or were dependent on N fixation. Nickel deprivation resulted in delayed nodulation and in a reduction of early growth. The addition of Ni (1 microgram/liter) to the nutrient media prevented urea accumulation, necrosis, and growth reductions. Extrapolating these findings, it is suggested that Ni may be essential for other higher plants. See Eskew reference listed for further detail.

Toxicity. Nickel contact dermatitis can occur among wearers of nickel-containing jewelry, more common among females than males. This is particularly true of nickel sulfate present in some jewelry. Localization of sites unexpectedly involve the ear lobes, neck, fingers, and wrists. Nickel is a major offender in connection with AECD (allergic eczematous contact dermatitis).

As mentioned earlier, nickel carbonyl is a volatile intermediate in the Mond process for nickel refining. This compound also is used for vapor plating of nickel in the semiconductor industry, and as a catalyst in the chemical and petrochemical industries. The toxicity of the compound has been known for many years. Exposure of laboratory animals to the compound has induced a number of ocular anomolies, including anophthalmia and microphthalmia, and has been shown to be a carcinogenic for rats (Sunderman et al., 1979).

NICKELINE. A nickel arsenide mineral, NiAs, crystallizes in the hexagonal system but is usually found massive. Color, light copper; hardness, 5.0–5.5; specific gravity, 7.784; luster, metallic; opaque. Found in several European localities and in the Province of Ontario, Canada; in the United States at Franklin, New Jersey, and Silver Cliff, Colorado. It is an ore of nickel.

NICOL PRISM. See **Prism (Optics).**

NICOTINAMIDE. See **Pyridine and Derivatives; Vitamins.**

NICOTINE. See **Alkaloids.**

NICOTINIC ACID. See **Pyridine and Derivatives; Vitamins.**

NIGHT BLINDNESS. See **Vision and the Eye.**

NIGHTINGALE (*Aves, Passeriformes*). A warbler, *Luscinia megarhyncha*, of western Europe, noted for its song. Farther east two other species, the eastern, *L. pheilomella*, and Persian, *L. hafizi*, nightingales, are found.

The nightingale is a trim, proud-looking bird, about 6 to 7 inches (15 to 18 centimeters) in length. The coloring is brown with lighter brown underneath. The female does the nesting and brooding, although the male helps with feeding the young. Incubation period is from 13 to 14 days. The eggs are blue or off-white with some markings.

NIGHTJARS AND NIGHTHAWKS (*Aves, Caprimulgiformes*). These birds make up the majority of the *Caprimulgiformes*, an order of nocturnal birds with very wide mouths. Nightjars have mottled plumage and a short beak. They are insect-eaters, flying chiefly at night. All continents have some of the numerous species with the exception of Australia. In North America, the nighthawk and whippoorwill are the most widely known representatives of the group, with the poorwill, chuck-will's widow (*Antrostomus carolinensis*), and the Merrill parauque (*Nyctidromus albicollus*) as less widely distributed species. Nightjars have been known since the time of Aristotle and are mentioned in the Bible. The poorwill (*Phalaenoplilus nuttallii*) is about 8 to 10 inches (20 to 25 centimeters) long and makes no formal nest. Eggs are laid in grasslands. This species uses the tactic of displaying a "broken" wing to distract the attention of predators. The legs are short and weak. The species is known for squatting lengthwise on limbs of trees and also for going into hibernation. During such periods, respiration is barely detectable and the temperature drops from a normal figure of about 100°F to 66°F (38°C to 18.9°C).

The whippoorwill, shown in Fig. 1, is a nocturnal bird (*Caprimulgus* (*Antrostomus*) *vociferus*), with a short beak and wide mouth, adapted

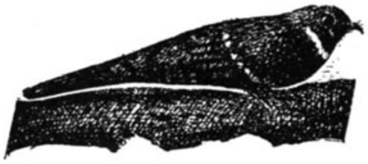


Fig. 1. Whippoorwill.

for taking insects in flight. Its call has been likened to the words used in its name. Often the three syllables are repeated over and over scores of times without cessation.

The common nighthawk (*Chordeiles minor* (*virginianus*)) is widely distributed in North America and winters far into South America. A second species, the Texan nighthawk (*C. acutipennis*) enters the southwestern United States. Nighthawks are characterized by very short beaks and wide mouths, well adapted for catching insects in flight. They commence their flights late in the day. Their flight is easy and powerful and their long dives, terminating in a peculiar hollow boom, are a memorable exhibition. See Fig. 2.

Fig. 2. Nighthawk.

Goatsuckers were given their odd name in the mistaken belief during Aristotle's time that the birds took milk from domesticated goats. They, like the other nightjars, are characterized by weak legs, a short weak beak, a very wide mouth, and crepuscular habits. The *Uropsalis lyra* is a small bird about 4 inches in length (10 centimeters), but with a lyre-type tail some 27 inches (69 centimeters) in length. This species inhabits the environs of Colombia in South America. The habits are nocturnal. The call is penetrating. Another species is the parauque, a large bird of Mexico and Texas which resembles the poorwill. Also related to the goatsuckers is the frog mouth (*Podargus*) found in the Oriental and Australian regions. See also **Caprimulgiformes.**

NIMBOSTRATUS. See **Clouds and Cloud Formation.**

NIMBUS. See **Clouds and Cloud Formation.**

NIOBIUM. Chemical element symbol Nb, at. no. 41, at. wt. 92.906, periodic table group 5, mp 2,458–2,468°C, bp 4,742°C, density 8.6 g/cm^3 (20°C). Elemental niobium has a body-centered cubic crystal structure. The metal has a slightly bluish tinge, is ductile and malleable, and when polished resembles platinum. The metal burns upon being heated in air. There is one natural isotope ^{93}Nb. Seven radioactive isotopes have been identified ^{90}Nb through ^{92}Nb and ^{94}Nb through ^{97}Nb, with a wide range of half-lives. ^{94}Nb has the longest half-life (2×10^4 years). The element was first identified by C. Hatchett in 1801 and was originally called columbium which name persisted for many years. The name still appears widely in the literature, particularly in connection with alloys bearing the element, such as columbium steels.

First ionizayion potential 6.77 eV; second 13.895 eV; third 24.2 eV. Oxidation potential $Nb \rightarrow Nb^{3+} + 3e^-$, ca. 1.1 V; $2Nb + 5H_2O \rightarrow Nb_2O_5 + 10H^+ + 10e^-$, 0.62 V.

Other important physical properties of niobium are given in the accompanying table and under **Chemical Elements.**

Niobium occurs, usually with tantalum, in columbite $Fe(NbO_3)_2$, (80% Nb_2O_5), pyrochlore (50% Nb_2O_5), samarskite (50% Nb_2O_5), chiefly found in western Australia, and South Dakota. Recovered along with tantalum by fusion with potassium bisulfate, and obtained in the residue after subsequent extraction with H_2O. Niobium and tantalum are separated by fractional crystallization of the potassium fluorides, niobium concentrating in the mother liquid and tantalum in the crystals.

The principal uses for the element are in alloys. Niobium also has gained prominence in research as a superconducting material. At the temperatures of liquid helium, niobium becomes a superconductor and, in the form of a fine wire, has been incorporated in a superconducting cell. The element has both size and cost advantages over electronic materials. The alloy Nb_3Sn becomes superconducting at a somewhat higher temperature. Niobium-titanium and niobium-zirconium alloys also have potential as superconductors.

Alloys. Niobium is used in steel, notably stainless steels, to stabilize the carbon present (as carbide) and for preparing niobium carbide, used for dies and cutting tools. Ferroniobium is a strong carbide-forming material and, when added to 18-8 stainless steel, stabilizes areas that may be heat-affected during welding and thus cause subsequent intergrantular corrosion. Niobium steels are used for rotors in gas turbines where temperatures up to 700°C must be withstood. Niobium-base alloys find application in fast reactors. Superalloys for very demanding use, as in military applications contain niobium with cobalt and zirconium. When alloyed with titanium, molybdenum, and tungsten, the elevated-temperature hardness of niobium in enhanced, whereas when alloyed with vanadium and zirconium, the strength of niobium up to temperatures of 500°C is increased. Metallurgically, niobium is attractive because of its density, good workability, retention of tensile strength at high temperatures, and its high melting point. In the temperature range 920–1,200°C, niobium has been found superior to most other metals on a strength-to-weight basis for aerospace applications. In multicomponent alloys, zirconium and hafnium when added with niobium add effectively to strength, even more so than molybdenum or tungsten, but there is some sacriffice in ductility.

In metallurgy, niobium is classified as a refractory metal, along with tungsten, tantalum, and molybdenum. A comparison of the four metals is given in the accompanying table.

Niobium in Tool Steels. In the matrix method of tool-steel development, the composition of the heat-treated matrix determines the steel's initial composition. Carbide volume-fraction requirements then are calculated, based upon historical data, and the carbon content is adjusted accordingly. This approach has been used to design new steels in which niobium is substituted for all or part of the vanadium present as carbides in the heat-treated material. Niobium provides dispersion hardening and grain refinement, and forms carbides that are as hard as vanadium, tungsten, and molybdenum carbides.

Chemistry and Compounds. Elemental niobium is insoluble in HCl or HNO_3, but soluble in hydrofluoric acid or a mixture of hydrofluoric and HNO_3.

As might be expected from its $4d^45s^1$ electron configuration, niobium forms pentavalent compounds. However, the stability of its compounds of lower valence is greater than that of the corresponding tantalum compounds, in keeping with the group 5 position of niobium and tantalum. Nevertheless the similarity of the properties of the compounds of the two metals is so great that special methods are required for their separation, such as solvent extraction of the pentachlorides or chromatographic removal of adsorbed TaF_5 with an ethylmethyl ketone-water system. In addition, divalent and tetravalent compounds are known, and an interstitial, nonstoichiometric hydride.

Niobium forms a divalent oxide, NbO, insoluble in water, but readily soluble in acids or NH_4OH. It also gives by direct combination of the metal on heating with oxygen, the pentoxide, Nb_2O_5, which can be reduced by hydrogen at high temperature to NbO_2, and on heating with magnesium to Nb_2O_3.

Niobium(III) halides are known, notably the chloride, $NiCl_3$, which is of particular interest because its solution has been shown to contain Nb^{3+} ions (in equilibrium with $NbCl_6^{3-}$ complex ions).

Tetravalent niobium is believed to occur in the form of $NbOCl_4^{2-}$ ions in a solution obtained, with color change, by reduction of HCl solution of $NbCl_5$, and by inference in similarly reduced solutions of the other pentahalides. Tetravalent niobium also is found in the dioxide (see above) and the carbide, NbC.

Four pentahalides of niobium, NbF_5, $NbCl_5$, $NbBr_5$, and NbI_5 have been prepared by heating the pentoxide with carbon in a current of the halogen. They are hydrolyzed in H_2O, and even in concentrated aqueous solution of the respective halogen acids; the Nb^{5+} ion is apparently not

REPRESENTATIVE PROPERTIES OF REFRACTORY METALS

Property	Tungsten	Tantalum	Molybdenum	Niobium
Density, g/cm^3	19.3	16.6	10.2	8.7
Melting point, °C	3,390–3,420	2,996	2,617	2,458–2,468
Boiling point, °C	5,660	5,325–5,525	4,612	4,742
Linear coefficient of expansion per °C	4.3×10^{-6}	6.5×10^{-6}	4.9×10^{-6}	7.2×10^{-6}
Thermal conductivity, 20°C ($cal/cm^2/cm/°C/s$)	0.40	0.13	0.35	0.13
Specific heat, 20°C (cal/g/°C)	0.032	0.036	0.061	0.065
Working temperature, °C	1,700	ambient	1,600	ambient
Electrical conductivity, % IACS	31	13	30	12
Nuclear cross section (thermal neutrons, Barns/atom)	19.2	21.3	2.4	1.1
Tensile strength, 1000 psi				
20°C	100–500	100–150	120–200	75–150
500°C	175–200	35–45	35–65	35
1,000°C	50–75	15–20	20–30	13–17
Young's Modulus of Elasticity, psi				
20°C	59×10^6	27×10^6	46×10^6	14×10^6
500°C	55×10^6	25×10^6	41×10^6	7×10^6
1,000°C	50 × 106	22 × 106	39×10^6	—
Poisson's Ratio	0.284	0.35	0.32	0.38
Corrosion resistance, 100°C				
Dilute HNO_3	See Tungsten	N	R	N
Dilute H_2SO_4		N	S	VS
Concentrated H_2SO_4		N	S	R
Dilute HCl		N	S	—
Concentrated HCl		N	SL	SL
Concentrated Hydrofluoric acid		R	SL	R
Phosphoric acid, 85%		N	SL	VS
Concentrated NaOH		R	N	R

N = no appreciable corrosion.
VS = <0.0005 inch (0.013 millimeter) per year.
SL = 0.0005–0.005 inch (0.013–0.13 millimeter) per year.
S = 0.005–0.01 inch (0.13 0.25 millimeter) per year.
R = >0.01 inch (0.25 millimeter) per year.

present, but rather complex ions such as $[NbOCl_4]^-$ or $[NbOCl_5]^{2-}$. The products of partial hydrolysis of the pentahalides are oxyhalides, such as $NbOF_3$, $NbOCl_3$, and $NbOBr_3$. They are designated in the older literature as columbyl or columboxy compounds. The more stable oxyhalogen compounds of niobium are complexes, such as $NbOF_3 \cdot 3NaF$, $NbOF_3 \cdot ZnO \cdot 6H_2O$, and $NbOF_3 \cdot 2KF \cdot H_2O$.

Further complexes of Nb(V) are formed with oxygen-function compounds, such as *o*-dihydroxybenzene and acetylacetone.

The so-called niobic acid is the hydrated pentoxide, $Nb_2O_5 \cdot xH_2O$, insoluble in H_2O.

The metaniobates of the alkali metals, $MNbO_3$, the orthoniobates M_3NbO_4 and the pyroniobates, $M_4Nb_2O_7$, where M is an alkali metal, can be prepared by various alkali carbonate or hydroxide fusion processes.

Niobium forms a nitride, NbN, and a carbide, NbC.

Niobium forms a diamino compound, $(NH_2)_2NbCl_3$, and an ammine complex, $NbCl_5 \cdot 9NH_3$. It forms two cyclopentadienyl compounds, $(C_5H_5)_2NbBr_3$ and $(C_5H_5)Nb(OH)Br_2$. Its other organometallic compounds are essentially oxygen-functional ones, such as $Nb(OCH)_3)_5$, $Nb(OC_2H_5)_5$, $Nb(O)(OC_5H_{11})_3$, and $Nb(OC_5H_{11})_5$. These compounds are named as substituted niobanes (thus, the last is pentabutoxy niobane) or as alkyl niobate esters.

Additional Reading

Carter, G. F., and D. E. Paul: "Materials Science and Engineering," ASM International, Materials Park, Ohio, 1991.

Staff: "ASM Handbook—Properties and Selection: Nonferrous Alloys and Pure Metals," ASM International, Materials Park, Ohio, 1990.

Staff: "Tool-Steel Developers Take Note of Niobium," *Advanced Materials 7 Processes*, 15 (June 1991).

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

Titran, R. H.: "Niobium and Its Alloys," *Advanced Materials & Processes*, 34 (November 1992).

NIST. The National Institute of Standards and Technology, the headquarters of which is located in Gaithersburg, Maryland, 20899. NIST replaces the former National Bureau of Standards (NBS), which was established by the U.S. Congress in 1901, with the objectives of (1) serving as the basis for the nation's physical measurement system, (2) providing scientific and technological services for industry and government, (3) establishing a technical basis for equity in trade, and (4) providing technical services to promote public safety.

As of 1993, NIST is comprised of several divisions and departments, including:

- Technology services
- Manufacturing technology centers
- Standards services
- Standards Code and information
- Standards management
- Weights and measures
- Laboratory accreditation
- Measurement services
- Standard reference materials
- Physical measurement services
- Research and technology applications
- Technology development and small business
- National technology workshop
- Information services
- Research resources development
- Research information services

NITER. This potassium nitrate mineral KNO_3 of orthorhombic crystallization usually occurs as thin crusts, or as silky acicular crystals. It has a hardness of 2, and specific gravity of 2.09–2.14, is of white color, translucent with vitreous luster. It occurs as a surface efflorescence, or in soils rich in organic material in arid regions. World occurrences include Spain, Italy, Egypt, Arabia, India, Russia and the western United States. Also in the Republic of South Africa, and Bolivia, South America. Large quantities were recovered from limestone caves in Tennessee, Kentucky, Alabama and Ohio during the Civil War for use in the manufacture of gunpowder. It is used as a source of nitrogen compounds, for explosives and fertilizers.

NITRATION. The process of adding nitrogen to a carbon compound, generally to create a nitro-derivative (adding a—NO_2 group) is termed nitration. An example is the formulation of nitrobenzene from benzene: $C_6H_6 + HNO_3 \rightarrow C_6H_5NO + H_2O$. In most instances, the—NO_2 group replaces a hydrogen atom. More than one hydrogen atom may be replaced, but each succeeding hydrogen represents a more difficult substitution. The nitrogen-bearing reactant may be: (1) strong HNO_3, (2) mixed HNO_3 and H_2SO_4, (3) a nitrate plus H_2SO_4, (4) nitrogen pentoxide N_2O_5, or (5) a nitrate plus acetic acid. Both straight-chain and ring-type carbon compounds can be nitrated. The alkanes yield nitroparaffins.

Various rules of addition govern the position of the entering nitro group, depending upon the conditions. For example, in the nonsubstituted benzene series, the nitro group can enter in the ortho, meta or para position, but the presence of some other group usually fixes the position of the entering nitro group. For example, it enters meta to a nitro, sulfonic, or carbonyl group, and ortho and para to a chloro, bromo, or hydroxy group. (These statements apply to the principal product formed, since in most substituted benzene reactions, a limited quantity of all ring positions are entered.) Various other rules govern other conditions in other aromatic series.

One of the great uses of nitration is to break into a pure hydrocarbon, which is usually more difficult to do by other means. The nitro group may then be changed and another group take its place. Typical examples are nitration of ethane to form nitroethane, and of benzene to form nitrobenzene, which is easily changed to aniline.

An important economic consideration in any nitration process is the recovery of the spent acid. Since the nitration reaction forms H_2O, the reagents gradually become diluted to a point where they will not react any more. The water may be taken up during the reaction by removing it with oleum or acetic anhydride, a practice which still leaves large amounts of the reagents at the end of the process.

The HNO_3 is usually concentrated by distilling it from H_2SO_4 solution which retains the H_2O. After the HNO_3 has been driven off, the temperature is raised and the H_2O is driven off the H_2SO_4, thereby concentrating the latter.

As an example of nitration, let us consider the preparation of nitrobenzene. Mixed acid consisting of strong H_2SO_4 plus HNO_3 is slowly added to benzene in a closed iron vessel provided with stirrer and reflux condenser. The acid must be added to the benzene. If it were done the other way, the benzene which was added first would be quickly nitrated all the way to a trinitrobenzene. The temperature is maintained from 45–55°C. After the nitration is finished, the nitro compound is separated from the acid by decantation, since the nitrobenzene is lighter and does not mix with the acid. The nitrobenzene is washed with water and with dilute caustic or sodium carbonate solution and then again with water to give a neutral product. To obtain dinitrobenzene the reaction would be run with stronger acid and at a temperature of about 100°C.

Since the reaction used concentrated H_2SO_4, ordinary iron vessels can be used, but the neutralization process must be carried out in lead-lined tanks. Good agitation and adequate cooling facilities are necessary to avoid any local overheating and the formation of higher nitrated compounds.

NITRIC ACID. This important industrial chemical has been known for at least 1000 years. The acid was known to alchemists as *aqua fortis* (strong water) or *aqua valens* (powerful water). Nitric acid was of particular interest to the early experimenters because of its ability to dissolve a number of metals, including copper and silver. Early chemists were also fascinated by the fact that addition of sal ammoniac (ammonium chloride) gave *aqua regia* (royal water) which dissolves gold as well as silver.

Nitric acid is a colorless liquid, sp. gr. 1.503 (25°C), freezing point –41.6°C, and boiling point 86°C. The 100% acid is not entirely stable and must be prepared from its azeotrope (constant-boiling mixture) by distillation with concentrated sulfuric acid. Reagent grade HNO_3 is a water solution containing about 68% HNO_3 (weight). This strength corresponds to the constant-boiling mixture of the acid with water, which is 68.4% HNO_3 (weight) at atmospheric pressure and boils at 121.9°C. Nitric acid is completely miscible with water. It forms two solid hydrates, $HNO_3 \cdot H_2O$ and $HNO_3 \cdot 2H_2O$, with corresponding melting points of approximately –38 and –18.5°C. Nitric acid is a strong acid and a powerful oxidizer. In dilute solutions, it is almost completely ionized to H^+ and NO_3^- ions and behaves like a strong acid.

With organic compounds, HNO_3 may act as a nitrating agent, as an oxidizing agent, or simply as an acid. The classic example of nitration is its reaction with benzene or toluene in the presence of concentrated H_2SO_4 to form nitrobenzene or nitrotoluene (TNT). An example of oxidation properties is in the oxidation of cyclohexanol by HNO_3 to produce adipic acid, an intermediate of nylon. Behaving like an acid, it forms nitroglycerin by esterification of glycerol in the presence of concentrated sulfuric acid.

An interesting property of HNO_3 is its ability to passivate some metals, such as iron and aluminum. This property is of significant industrial importance, since modern processes for producing the acid depend on it. Modern suitability formulated stainless steel alloys are usefully resistant to nitric acid through a wide range of conditions. The acid's passivity or the metal's resistance to attack is attributed to the formation of a protective oxide layer on the surface of the metal.

Nitric acid is a high tonnage industrial chemical. Much of the production is used in the manufacture of agricultural fertilizers, largely in the form of ammonium nitrate, NH_4NO_3. See **Fertilizers.** About 15% of the nitric acid produced is used in explosives (nitrates and nitro compounds), and about 10% is consumed by the chemical industry. As the red fuming acid or as nitrogen tetroxide, HNO_3 is used extensively as the oxidizer in propellants for space rockets and missiles.

Production of Nitric Acid. Three commercial methods have been developed for nitric acid production: (1) the reaction between sulfuric acid and sodium nitrate, (2) the thermal combination of oxygen and nitrogen in air, and (3) the catalytic oxidation of ammonia and absorption of the gaseous products in waters. There are numerous variations of these fundamentals processes. The principal process used today is based on the catalytic oxidation of ammonia and absorption of the gaseous products in water. This process was developed by Ostwald (Germany) and based on earlier work of Kuhlmann (France). In the Ostwald process, HNO_3 is produced in a 3-stage operation: (1) Ammonia is oxidized to nitric oxide, (2) the nitric oxide is further oxidized to nitrogen dioxide, and (3) the gases are absorbed in water to yield HNO_3 according to

$$4NH_3 + 5O_2 \rightarrow 4NO + 6H_20$$
$$2NO + O_2 \rightarrow 2NO_2$$
$$3NO_2 + H_2O \rightarrow 2HNO_3 + NO$$

The nitric oxide formed in the last equation returns to the gas phase, is reoxidized to nitrogen dioxide, and reabsorbed. These reactions are highly exothermic. In actuality, numerous complex reactions occur in addition to the main reactions just outlined.

In a manufacturing plant, air is preheated, mixed with superheated ammonia vapor, and reacted catalytically over a gauze composed of 90% platinum and 10% rhodium at a temperature of 800–960°C and operating pressures between 1 and 8.2 atmospheres. The reaction produces nitrogen dioxide, NO_2 and nitric oxide, NO. The latter is oxidized to NO_2 in the reaction train. The NO_2 actually exists in equilibrium with its dimer, N_2O_4. This equilibrium mixture, sometimes referred to as nitrogen peroxide, is absorbed in water in a cooled absorber tower to form HNO_3 at a strength of 55–60% HNO_3.

NITRIDING. Surface hardening of alloy steels by heating the metals to a temperature of 490–650°C in an atmosphere of partially dissociated

NH_3 (ammonia). As in cyaniding, hardening results from the formation of nitrides of iron and of certain alloying elements that may be present in the steel. Much longer heating time is required than in carburizing practice, and while the depth of penetration is generally less, the maximum hardness at the surface is higher, 900–1,100 D.P.H. (*Vickers* Brinell) compared to 800–900 D.P.H. for an average carburized case. Nitriding also differs from carburizing in that the parts are fully heat-treated to develop the required core properties before the nitriding treatment. Because of the comparatively low temperature of the process, distortion and dimensional changes are at a minimum. Nitrided steels have good corrosion-resistance when used for valves, pump parts, shafting, and bearing surfaces operating in steam, crude oil, gasolines, and gaseous products of combustion. The fatigue strength is also improved by nitriding.

Other typical applications are piston pins, crankshafts, cylinder liners, timing gears, gauges, and ball and roller bearing parts.

NITRILE RUBBER. See **Elastomers.**

NITRILES. See **Amines.**

NITRO- AND NITROSO-COMPOUNDS. Nitro-compounds contain the nitro-group (—NO_2) attached directly to a carbon atom; nitroso-compounds contain the nitroso-group (—NO) similarly attached. A very important member of this group is nitrobenzene, which upon reduction yields a variety of products, important in the synthesis of drugs and dyes. See accompanying table.

Alkylnitro-Compounds:

Primary	Secondary	Tertiary
$CH_3CH_2{\cdot}NO_2$	$(CH_3)_2CH{\cdot}NO_2$	$(CH_3)_3C{\cdot}NO_2$
Nitroethane	Nitrodimethylmethane (2-nitropropane)	Nitrotrimethylmethane

Isomeric Nitrites:

$CH_3CH_2{\cdot}ONO$	$(CH_3)_2CH{\cdot}ONO$	$(CH_3)_3{\cdot}ONO$
Ethyl nitrite	Isopropyl nitrite	1,1-dimethylethyl nitrite

Alkylnitroso-Compounds:

$(CH_3)_3C{\cdot}NO$
Nitrosotrimethylmethane

Nitrates:

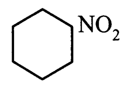
$CH_3CH_2{\cdot}ONO_2$	$(CH_3)_2CH{\cdot}ONO_2$	$(CH_3)_3C{\cdot}ONO_2$
Ethyl nitrate	Isopropyl nitrate	1,1-dimethylethyl nitrate

Nitrosamine:

$(C_2H_5)_2N{:}NO$
Diethylnitrosamine

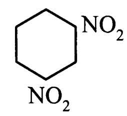
Benzenoid Nitro- and Nitroso-Compounds:

Mononitro-compound



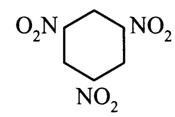
Nitrobenzene

Dinitro-compound



1,3-Dinitrobenzene

Trinitro-compound



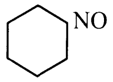
1,3,5-Trinitrobenzene

REPRESENTATIVE NITRO- AND NITROSO-COMPOUNDS

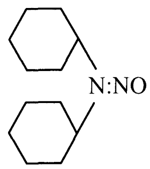
Compound	Formula	Melting Point, °C	Boiling Point, °C
	REPRESENTATIVE NITRO-COMPOUNDS		
Nitrobenzene	$C_6H_5{\cdot}NO_2$	6	211
1,3-Dinitrobenzene	$C_6H_4(NO_2)_2$ (1,3)	90	302
2-Nitrotoluene	$CH_3C_6H_4(NO_2)$ (2)	−11	222
2,4-Dinitrotoluene	$CH_3C_6H_3(NO_2)_2$ (2,4)	70	300
Trinitrotoluene ("T.N.T.")	$CH_3C_6H_2(NO_2)_3$ (2,4,6)	81	240 expl.
3-Nitrophenol	$HOC_6H_4{\cdot}NO_2$ (3)	96	194 (70 torr)
2,4,6-Trinitrophenol (picric acid)	$HOC_6H_2(NO_2)_3$ (2,4,6)	122 expl.	>300
4-Nitrobenzaldehyde	$C_6H_4(CHO)(NO_2)$ (1,4)	58	164 (23 torr)
4-Nitrobenzoic acid	$C_6H_4(COOH)(NO_2)$ (1,4)	240	subl.
4-Nitrobenzyl alcohol	$C_6H_4(CH_2OH)(NO_2)$ (1,4)	93	185 (12 torr)
2-Nitronaphthalene	$C_{10}H_7(NO_2)$ (2)	79	165 (15 torr)
1-Nitroanthraquinone	$C_6H_4(CO)_2C_6H_3(NO_2)$ (1)	230	subl.
2-Nitropropane	$(CH_3)_2CHNO_2$	−93	120
Nitroethyl alcohol	$CH_2OHCH_2NO_2$	<−80	194
Nitrobromoform (bromopicrin)	NO_2CBr_3	10	expl.
Nitrochloroform (chloropicrin)	NO_2CCl_3	−64	112
Nitrofurane	$C_4H_3O{\cdot}NO_2$	28	
Nitrourea	$OC\langle {NH_2 \atop NHNO_2}$	155 dec.	
Nitroguanidine	$HNC\langle {NH_2 \atop NHNO_2}$	246	
1,3-Nitroaniline	$C_6H_4(NO_2)(NH_2)$ (1,3)	114	>285
	REPRESENTATIVE NITROSO-COMPOUNDS		
Nitrosobenzene	C_6H_4NO	68	58 (18 torr)
4-Nitrosophenol (4-quinoneoxime)	$C_6H_4(OH)(NO)$ (1,4)	125	144 dec.
4-Nitrosonaphthol-1 (4-naphthaquinoneoxime)	$C_{10}H_6(OH)(NO)$ (1,4) or $C_{10}H_6(O)(NOH)$ (1,4)	193	
2-Nitrosonaphthol-1	$C_{10}H_6(OH)(NO)$ (1,2)	163 dec.	
N-Nitrosomethylaniline	$C_6H_5N\langle {CH_3 \atop NO}$	13	128 (20 torr)
4-Nitrosophenylaniline	$C_6H_5NH{\cdot}C_6H_4NO$	145	
1-Nitrosonaphthylamine-2	$C_{10}H_6(NH_2)(NO)$ (2,1)	151	
Diphenylnitrosamine	$(C_6H_5)_2N{\cdot}NO$	66	

dec., decomposes; expl., explodes; subl., sublimes

Nitroso-compounds



Nitrosobenzene



Diphenylnitrosamine

Under the proper conditions of concentration of HNO_3 and of temperature, benzene forms mainly nitrobenzene, nitrobenzene forms mainly 1,3-dinitrobenzene, and 1,3-dinitrobenzene, mainly 1,3,5-trinitrobenzene.

When nitrobenzene is treated (1) with zinc and calcium chloride or ammonium chloride solution, beta-phenylhydroxylamine, C_6H_5NHOH, is formed, and from this by treatment with chromic acid or ferric chloride nitrosobenzene is formed, (2) with tin or iron and HCl, aniline, $C_6H_5NH_2$, is formed and from this by treatment with nitrous acid followed by treatment with stannous chloride plus HCl phenylhydrazine, $C_6H_5NH{\cdot}NH_2$, is formed.

Mono- or poly-substituted nitro-compounds are changed in whole or in part to the corresponding amino-compounds by proper choice of reducing agent and temperature, e.g., in acid medium 1,3-dinitrobenzene yields 1,3-phenylenediamine, $C_6H_4(NH_2)_2(1,3)$, and with ammonium sulfide yields 3-nitroaniline $(1)H_2NC_6H_4NO_2(3)$. When diphenylnitrosamine is reduced, 1,1-diphenylhydrazine, $(C_6H_5)_2N{\cdot}NH_2$, is formed.

See also **Nitration.**

NITROCELLULOSE. See **Cellulose.**

NITROGEN. Chemical element, symbol N, at. no. 7, at. wt. 14.0067, periodic table group 15, mp −209.86°C, bp −195.8°C, critical temperature −147.1°C, critical pressure 33.5 atmospheres, density 1.14 g/cm^3 (solid), 1.25057 g/L (0°C, 760 torr), 0.9675 (air = 1.0000). Solid nitrogen has a hexagonal crystal structure. Nitrogen at standard conditions is a colorless, odorless, tasteless gas. The gas is slightly soluble in H_2O (2.35 parts nitrogen in 100 parts H_2O at 0°C), the solubility decreasing with increasing temperature (1.55 parts nitrogen in 100 parts H_2O at 20°C). Nitrogen is slightly soluble in alcohol and is essentially insoluble in most other known liquids. There are two naturally occurring isotopes, ^{14}N and ^{15}N, with ^{14}N by far the most abundant (99.635%). Four radioactive isotopes have been identified, ^{12}N, ^{13}N, ^{16}N, and ^{17}N, all with extremely short half-lives measured in seconds or minutes. In terms of abundance in igneous rocks in the earth's crust, nitrogen does not appear among the first 37 most abundant elements. In terms of abundance in seawater, nitrogen ranks 16th, with an estimated 2,300 tons of nitrogen per cubic mile of seawater. In terms of cosmic abundance, nitrogen ranks 7th. For comparison, assigning a value of 10,000 to silicon, the figure for nitrogen is 160,000 and that for hydrogen, estimated the most abundant, a figure of 3.5×10^8. Of dry air in the earth's atmosphere, disregarding pollutants, 78.09% is nitrogen by volume and 75.54% by weight. In the atmosphere, the nitrogen is mixed with oxygen, argon, the rare gases, CO_2, and H_2O vapor. Nitrogen was first identified as an element by Daniel Rutherford in 1772. Lavoisier further confirmed Rutherford's findings in 1776. Like oxygen, nitrogen is essential to practically all forms of life, making some of the compounds of this element extremely important as foods and fertilizers. Nitrogen serves the important function of diluent in the earth's atmosphere, controlling natural burning and respiration rates that otherwise would proceed much faster with higher concentrations of oxygen. Nitrogen is an important ingredient of numerous inorganic and organic compounds, including alkaloids, amides, amines, cyanides, cyanogens, diazo compounds, hydrazines, imides, nitrates, nitrides, nitrites, nitriles, oximes, purines, pyridines, and ureas. In terms of high-tonnage production, the nitrogen compound NH_3 (ammonia) ranks first with worldwide production exceeding 50 million tons annually.

First ionization potential 14.84 eV; second, 29.47 eV; third, 47.17 eV; fourth, 73.5 eV; fifth, 97.4 eV. Oxidation potentials $H_2N_2O_2 + 2H_2O \rightarrow 2HNO_2 + 4H^+ + 4e^-$, −0.80 V; $N_2O_4 + 2H_2O \rightarrow 2NO_3^- + 4H^+ + 2e^-$, −0.81 V; $HNO_2 + H_2O \rightarrow NO_3^- + 3H^+ + 2e^-$, −0.94 V; $NO + 2H_2O \rightarrow NO_3^- + 4H^+ + 3e^-$, −0.96 V; $NO + H_2O \rightarrow HNO_2 + H^+ + e^-$, −0.99 V; $2NO + 2H_20 \rightarrow N_2O_4 + 4H^+ + 4e^-$, −1.03 V; $2HNO_2 \rightarrow N_2O_4 + 2H^+ + 2e^-$, −1.07 V. $N_2O + 3H_2O \rightarrow 2HNO_2 + 4H^+ + 4e^-$, −1.29 V; $N_2O + H_2O \rightarrow 2NO + 2H^+ + 2e^-$, −1.59 V; $N_2 + H_2O \rightarrow N_2O + 2H^+ + 2e^-$, −1.77 V; $N_2O_4 + 4OH^- \rightarrow 2NO_3^- + 2H_2O + 2e^-$, 0.85 V; $NO + 2OH^- \rightarrow NO_2^- + H_2O + e^-$, 0.46 V; $N_2O_2^{2-} + 4OH^- \rightarrow 2NO_2^- + 2H_2O + 4e^-$, 0.18 V; $NO_2^- + 2OH^- \rightarrow NO_3^- + H_2O + 2e^-$, −0.01 V; $N_2O_2^{2-} \rightarrow 2NO + 2e^-$, −0.10 V; $N_2O + 6OH^- \rightarrow 2NO_2^- + 3H_2O + 4e^-$, −0.15 V; $N_2O + 2OH^- \rightarrow 2NO + H_2O + 2e^-$, −0.76 V. $2NO_2^- \rightarrow N_2O_4 + 2e^-$, −0.88 V.

Other physical properties of nitrogen are given under **Chemical Elements.**

Industrial Nitrogen

Like many of the elements, the compounds of nitrogen by far exceed the use of elemental nitrogen (discounting its important role as diluent in the atmosphere). Industrially, nitrogen gas is produced as a by-product in the liquefaction of air to produce pure oxygen. For some applications, nitrogen provides an excellent inert atmosphere for electric furnace operations and for the gaseous insulation of transformers. An inert atmosphere is required where air must be excluded. Nitrogen is one of the three main gases used for such atmospheres, the other two being carbon monoxide and hydrogen. In providing an inert atmosphere, nitrogen reduces the velocities of reactions, lowers the partial pressure and reduces the flammability of any active gases that may be present. Since commercial nitrogen usually contains traces of oxygen, H_2O vapor, and CO_2, sufficient to cause some oxidation at high temperatures, methane may be added to make the gas fully inert.

Nitrogen gas also is required for nitriding certain alloy steels, but pure gas is not required. The nitrogen is provided by dissociating ammonia at the process temperatures ranging from 475–650°C. Metals treated in this manner are hardened by the formation of nitrides on their surface (casehardening). In cyaniding, iron-base alloys simultaneously absorb carbon and nitrogen by heating the metals in a cyanide salt. Again, the nitrogen is not required in initial gaseous form. See also **Nitriding.** Several powder metallurgy techniques also utilize dissociated NH_3 atmospheres.

Environmental Aspects of Nitrogen

The oxides of nitrogen are among the most critical of air pollutants—both in their effects and in their abatement. These aspects of nitrogen are discussed under **Pollution (Air).**

Chemistry and Compounds

Most of the high-tonnage nitrogen-bearing compounds are described elsewhere in this volume. See also **Ammonia; Ammonium Chloride; Ammonium Hydroxide; Ammonium Nitrate; Ammonium Phosphates; Ammonium Sulfate;** and **Fertilizer.**

In the laboratory, nitrogen, mixed with argon, neon, krypton, and xenon, is obtained from the air by passing it over heated copper to remove the oxygen, or pure by fractional distillation of liquid air whereby the nitrogen distills off before the oxygen. Pure nitrogen may also be obtained by heating such compounds as ammonium nitrite and ammonium dichromate, and collecting the gas. Mixed with carbon monoxide in producer gas, nitrogen may be utilized without separation by first making methyl alcohol from carbon monoxide and hydrogen and then using hydrogen and nitrogen for ammonia. When nitrogen at low pressure is subjected to the silent electric discharge, activated nitrogen is produced. Activated nitrogen displays a golden yellow afterglow upon cessation of the current, increased by cooling and decreased by heating. This form of nitrogen is very active with phosphorus, with alkali metals (forming azides), with the vapor of zinc, mercury, cadmium, arsenic (forming nitrides), with many metallic chlorides (forming a green fluorescence), and with hydrocarbons (forming hydrocyanic acid and cyanides). The transformation of nitrogen to activated nitrogen is partial, and its return to ordinary nitrogen takes place rapidly, in about one minute.

The metal amides and imides are important in the nitrogen system. The amides of the active metals are produced by (1) reaction of the metal with NH_3, (2) reaction of the metal hydride with NH_3, (3) reaction of the metal nitride with ammonia, (4) reaction with another amide, as

$KNH_2 + NaI \rightarrow NaNH_2 + KI$ (in liquid NH_3). This last method is generally useful for the preparation of the heavy metal amides and imides from halides and binary halogenoids of the heavy metals. Cadmium amide, $Cd(NH_2)_2$ and lead imide, PbNH, for example, are readily prepared in this way. In some cases neither the amide nor the imide is stable, and the reaction proceeds to the nitride.

$$3HgBr_2 + 6KNH_2 \xrightarrow[\text{liq}]{NH_3} Hg_3N_2 + 6KBr + 4NH_3$$

The metal amides and imides are very reactive with oxygen, and are often unstable or even explosive. Some nitrides (e.g., of silver, gold, and mercury) are explosive, but others are stable. The latter may be obtained, (1) by reaction with the metal with nitrogen or ammonia at higher temperatures, e.g., aluminum nitride and magnesium nitride, AlN and Mg_3N_2, (2) by deamination of the metal amide or azide on heating, e.g., Ba_3N_2. The great thermal stability of certain nitrides, e.g., those of boron, silicon and phosphorus, BN, Si_3N_4 and P_3N_5, is attributed to polymerization. Many of the transition metal nitrides are interstitial compounds and are hard and metal-like in their properties.

In the nitrogen system, hydrazine is analogous to hydrogen peroxide in the oxygen system, its structure being

$$\begin{array}{c} \text{H}\;\; \\ \text{H:}\ddot{\text{N}}\text{:}\ddot{\text{N}}\text{:H} \\ \;\;\text{H} \end{array}$$

It is readily oxidized, even undergoing auto-oxidation under many conditions, and it is a powerful reducing agent. Like hydrogen peroxide it readily disproportionates (e.g., with a platinum catalyst), giving nitrogen and NH_3. Its reactivity (and other properties) makes it, and its derivative, unsymmetrical dimethylhydrazine, important rocket fuels. It forms addition compounds with many substances, including a monohydrate with H_2O. Hydrazine ($pK_{B1} = 6.04$, $pK_{B2} = 14.88$) forms hydrazinium(1+) compounds, containing the $N_2H_5^+$ ion, analogous to ammonium, and hydrazinium(2+) compounds containing the $N_2H_6^{2+}$ ion.

Hydroxylamine is related in its structure both to hydrazine (see formula above) and to hydrogen peroxide.

$$\begin{array}{c} \text{H} \\ \text{H:}\ddot{\text{N}}\text{:}\ddot{\text{O}}\text{:} \\ \text{H} \end{array} \qquad \begin{array}{c} \text{H} \\ \text{:}\ddot{\text{O}}\text{:}\ddot{\text{O}}\text{:} \\ \text{H} \end{array}$$

The chemical properties of hydroxylamine also suggest a compound intermediate between hydrazine and hydrogen peroxide. Its bond lengths are, N—O, 1.46Å, N—H, 1.01Å, O—H, 0.96Å, and its angles are H—O—N, 103°, H—N—O, 105°, and H—N—H, 107°. It is a base ($pK_B = 9.02$), forming salts containing the hydroxylammonium ion $HONH_3^+$.

Hydrazoic acid, HN_3, $pK_A = 4.72$, and most of its covalent compounds (including its heavy metal salts) are explosive. It is formed (1) in 90% yield by reaction of sodium amide with nitrous oxide, (2) by reaction of hydrazinium ion with nitrous acid, (3) by oxidation of hydrazinium salts, (4) by reaction of hydrazinium hydrate with nitrogen trichloride (in benzene solution). Hydrazoic acid forms metal azides with the corresponding hydroxides and carbonates. It reacts with HCl to give ammonium chloride and nitrogen, with H_2SO_4 to form hydrazinium acid sulfate, with benzene to form aniline, and it enters into a number of oxidation-reduction reactions.

The azides, except those of mercury(I), Hg(I), thallium(I), Tl(I), copper, Cu, silver, Ag, and lead, Pb, are readily prepared from hydrazoic acid and the oxide or carbonate of the metal, or by metathesis of the metal sulfate with barium azide. They are all thermally unstable, giving nitrogen and free metal or occasionally nitride. The azide ion appears to resonate between four structures:

$$^{-}:\ddot{N}{=}\overset{+}{N}{=}\underset{..}{N}:^{-}\;,\quad ^{-}:\underset{..}{N}{=}\overset{+}{N}{=}\ddot{N}:^{-}$$

$$^{=}:\underset{..}{\ddot{N}}{-}\overset{+}{N}{\equiv}N:^{-}\;,\quad ^{-}:N{\equiv}\overset{+}{N}{-}\underset{..}{\ddot{N}}:^{=}$$

These structures are in accord with a spacing of 1.15Å and electronic charges of −0.83, 0.66, and 0.83 on the three nitrogen atoms.

N(I) compounds. Hydration of nitrogen(I) oxide, N_2O to hyponitrous acid, $H_2N_2O_2$, is not possible. However, the latter decomposes (in three steps) to yield the former, which is thus its anhydride. Spectroscopic studies indicate a linear structure for N_2O, resonating between

$$^{-}:\ddot{N}{=}\overset{+}{N}{=}\ddot{O}: \quad \text{and} \quad N{\equiv}\overset{+}{N}{-}\underset{..}{\ddot{O}}:^{-}$$

However, heat capacity measurements give a higher entropy at low temperatures than spectroscopic studies do, which is explained by a partial randomness of the structure at low temperatures.

Hyponitrous acid ($pK_{A1} = 7.05$, $pK_{A2} = 11.0$) and its salts are obtained by (1) reduction of sodium nitrite with (a) sodium amalgam, (b) by electrolysis, (c) by stannous or ferrous salts, (2) by reduction of alkyl nitrates, (3) by reduction of hydroxylamine by noble metal oxides, and (4) by reduction of sodium hydroxylamine monosulfonate in alkaline solution.

Explosive salts such as NaNO can be prepared by the reaction of NO and liquid ammonia solutions of alkali metals. The unstable free acid, HNO, is thought to be an intermediate in many redox reactions of nitrogen compounds.

Nitramide, NO_2NH_2, a weak acid ($pK_A = 6.59$), is relatively more stable than its isomer hyponitrous acid.

N(II) Compounds. Nitrogen(II) oxide is formed in many reductions of nitrous acid, but is best prepared pure by reduction with ferrous ions, Fe^{2+}, or iodide ions, I^-. It undergoes many types of addition reactions, but its very slight tendency to dimerize and its low reactivity under ordinary conditions suggest that its odd electron lies in an antibonding orbital of very low energy; and the molecular orbital formulation is

$$NO[KK(z\sigma)^2(y\sigma^*)^2(x\sigma)^2(w\pi)^4(v\pi^*)]$$

The nitrosyl compounds can be readily classified on the basis of three modes of reaction of the NO molecule in accordance with the above formulation.

1. It can lose (or partly lose) the odd electron to form an ion of the formula $:N{\equiv}O:^+$. This formula gives rise to ONF, ONCl and ONBr by direct reaction of NO and the halogen. These are covalent compounds. Such salts as $NOBF_4$, $NOPF_6$, $NOAuF_4$, $NOSO_3F$, and $NOHSO_4$, on the other hand, are ionic. These may be considered the salts of nitrous acid acting as a base, $ONOH \leftrightarrows NO^+ + OH^-$, $pK_B = 18.2$.
2. It can gain an electron to form a negative ion of the formula

   $$N{\equiv}\ddot{O}:^{-}$$

   Thus dry NO reacts with sodium in liquid ammonia to form sodium nitrosyl, NaNO (empirical formula).
3. It can share a pair of electrons to form a coordinate link, as it does in coordination compounds. In most of these, it appears to coordinate as the positive ion, by transfer of an electron to an acceptor metal, which is thereby reduced by 1 unit in oxidation state. This causes, in some cases, the need for placing a negative charge on the metal. To avoid this, Pauling assumed the presence of four bonding electrons, involving structures of the type

   $$M{=}\overset{+}{N}{=}\ddot{O}:$$

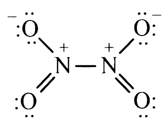
Nitrogen(III) Compounds. Nitrogen(II) oxide, NO, readily enters into equilibrium with NO_2 to form N_2O_3, nitrogen sesquioxide. The latter is unstable even at room temperature and consists of an equilibrium mixture of the three compounds. Its structure appears to be O═N—NO_2. If an equimolar mixture of NO and NO_2 is cooled and condensed, a blue liquid, bp 3.5°C, largely N_2O_3, is obtained. The latter readily combines with H_2O to form nitrous acid, HNO_2 ($pK_A = 3.29$). Nitrous acid is unstable, forming the equilibrium mixture, $3HNO_2 \leftrightarrows NO_3^- + 2NO + H_3O^+$, which in concentrated solution or on warming is largely displaced to the right (K = 39.6 at 30°C). Moreover, the NO undergoes further reactions, so that the actual system is complex. One of these reactions is: $NO + OH \rightarrow NO^+ + OH^- \leftrightarrows NO{\cdot}OH \leftrightarrows HNO_2$.

The existence of NO^+ and NO^- helps to explain the kinetics of nitrous acid as an oxidizing agent. It oxidizes I^-, Sn^{2+}, Fe^{2+}, Ti^{3+}, $S_2O_3^{2-}$, SO_2, and H_2S. It reacts with NH_3, urea, sulfonates and some other nitrogen compounds to produce nitrogen. With aromatic amines in the cold, it gives diazo compounds, while with secondary amines it gives nitroso compounds. Nitrous acid also functions as a reducing agent, as in the reactions with permanganate and hydrogen peroxide, in which nitrate ion is formed.

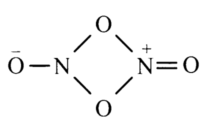
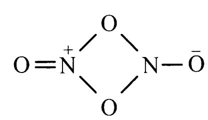
The nitrites vary widely in solubility, those of the alkalies and alkaline earths being very soluble, while those of the heavy metals are only slightly so. Moreover, the latter are relatively unstable, some decomposing at room temperature. The nitrites, like nitrous acid, function either as oxidizing or reducing agents. X-ray and spectroscopic studies give a triangular structure for the nitrite ion, with the N—O bond length 1.13Å and the O—N—O angle 120–130°. Values of 1.23Å and 116° have also been reported. Complex ions containing the NO_2 group may be either nitrito complexes (e.g., $Co(NH_3)ONO^{2+}$) or nitro complexes (e.g., $Co(NH_3)NO_2^{2+}$). The former of these two examples readily isomerizes to the latter.

Nitrosyl fluoride, NOF, and nitrosyl chloride, NOCl, are quite stable, but the bromide decomposes at room temperature. They are prepared by direct union of NO and the halogen, among other methods. Three trihalides, NF_3, NCl_3, and NI_3, are known. The first is a colorless stable gas; NCl_3 is a yellow liquid and NI_3, a brown solid; both are explosive. The contrast in stability is attributed to the large amount of ionic resonance energy of the N—F bond, which gives NF_3 a negative heat of formation.

Nitrogen(IV) Compounds. Nitrogen dioxide, NO_2, readily associates to form the tetroxide, N_2O_4, so that at ordinary temperatures and pressures both forms are present in equilibrium. Since nitrogen dioxide has an unpaired electron, it is paramagnetic and colored (red). N_2O_4 is diamagnetic and colorless. As with NO, the odd electron is in an antibonding orbital but of higher energy so that NO_2 is more reactive and more readily undergoes dimerization. The N—O bond length is 1.20Å and the angle is 132° (electron diffraction). The structure of N_2O_4 is, on the basis of spectral and entropy considerations,



This formula is at variance with Pauling's stability argument, but is supported by Ingold's evidence (*Nature*, **159**, 743, 1947). Longuet-Higgins has proposed the structures

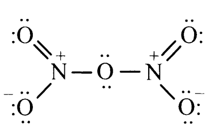
 and 

Nitrogen dioxide molecules react with NO to form N_2O_3, in an equilibrium mixture. The equilibrium mixture of NO_2 and N_2O_4 also reacts with water in a series of reactions

$$2NO_2 + H_2O \leftrightarrows H^+ + NO_3^- + HNO_2$$
$$3HNO_2 \leftrightarrows H^+ + NO_3^- + 2NO + H_2O$$

In warm solution, at high acidity, the second reaction is very rapid. In basic solutions the simple disproportionation $N_2O_4 + 2OH^- \rightarrow NO_2^- + H_2O$ takes place.

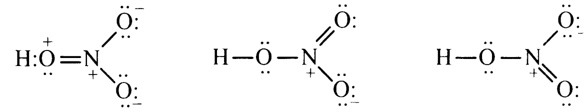
Nitrogen(V) Compounds. Nitrogen(V) oxide, N_2O_5, the anhydride of nitric acid, is a white solid subliming at 32.4°C and 760 mm. It hydrates readily to HNO_3, is a strong oxidizing agent, and decomposes at 20°C slowly into NO_2 and O_2. Its structure in the gas state consists of the molecules



However, x-ray, Raman, and infrared spectra show the crystalline solid to consist of NO_2^+ and NO_3^- ions.

Pure nitric acid, HNO_3, is a colorless liquid boiling with decomposition at 86°C and 760 torr. Upon continued heating it decomposes into NO_2, O_2 and H_2O. It is a fairly strong acid ($K_A = 22$), showing dissociation in concentrated solutions, and the presence of nitryl cation, NO_2^+ (nitronium ion). Solutions of HNO_3 in H_2SO_4 owe many of their properties to ions such as NO_2^+ and NO^+, as well, of course, as to HSO_4^- and oxonium ions.

The properties of HNO_3 are in accordance with resonance between the three electronic structures:

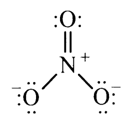


in which the last formula contributes a relatively small proportion to the overall structure. The two N—O bond lengths are 1.22Å, and N—O—H bond lengths 1.41Å and 0.96Å. The N—O—H angle is 90° and the O—N—O angle 130°.

The reactions of nitric acid are of three types: (a) acid-base reactions which are typical of a strong acid; (b) oxidation reactions, such as those with metals and organic materials, the latter often involving carbonization; (c) substitution reactions such as the replacement of —H by —NO_2 in aromatic hydrocarbons, to form nitro compounds, or of hydroxyl hydrogen by —NO_2 to produce esters of HNO_3.

These esters of nitric acid form one of the two groups of nitrates, the covalent group, which are also exemplified by nitryl hypofluorite and hypochlorite ($FONO_2$ and $ClONO_2$), often called fluorine and chlorine nitrate. Most nitrates, however, are ionic, e.g., salts of HNO_3. All metal nitrate are soluble in H_2O. Anhydrous metal nitrates, such as $Cu(NO_3)_2$. $Ti(NO_3)_4$, $VO(NO_3)_3$, $CrO_2(NO_3)_2$, $Si(NO_3)_4$, can be made by the action of liquid N_2O_4 on the metal (e.g., Cu) or of $ClONO_2$ on the corresponding chloride (e.g., the other examples given above).

The nitrate ion is considered to resonate between three equivalent structures of the form:



Two nitryl halides, NO_2F and NO_2Cl, are known, as well as nitryl salts, such as NO_2AsF_6, NO_2SbF_6, $(NO_2)_2SiF_6$, NO_2ClO_4, etc. Nitrogen also forms higher oxides, such as NO_3, and possibly NO_4, under action of the electric discharge.

Nitrate Losses from Disturbed Forest Ecosystems

Nutrient losses occur following a forest harvest or other disturbance, whether natural or anthropogenic. Studies have shown a variety of patterns of such losses. Vitousek et al. (1979) report on a systematic examination of nitrogen cycling in disturbed forest ecosystems and show that at least 8 processes, operating in 3 stages in the nitrogen cycle, can delay or prevent solution losses of nitrate from disturbed forests. The study involved 19 forest sites in the United States, including Pack Forest, Findley Lake, and Cascade Head in the northwest; Tesuque Watersheds in the southwest; Lake Monroe in southern Indiana; Coweeta in southwestern North Carolina; and Harvard Forest, Mount Mossilauke, and Cape Cod in the northeastern United States.

The 3 stages and 8 operative processes identified are:

Stage 1. Processes preventing or delaying ammonium accumulation.
(a) Nitrogen immobilization
(b) Ammonium fixation
(c) Ammonia volatilization
(d) Plant nitrogen uptake

Stage 2. Processes preventing or delaying nitrate accumulation.
(e) Lag in nitrification
(f) Denitrification to: $-N_2$, N_2O, or NO_x, $-NH_4$

Stage 3. Processes preventing or delaying nitrate mobility.
(g) Lack of water
(h) Nitrate sorption
(f) Denitrification at depth

The researchers stress that the net effect of all of these processes, except uptake by regrowing vegetation, is insufficient to prevent or delay losses from relatively fertile sites and thus such sites have the potential for very high nitrate losses following disturbance.

Nitrogen Fixation

A positive balance of usable nitrogen on earth depends upon nitrogen fixation which is the process by which atmospheric nitrogen, N_2, is converted either by biological or chemical means to a form of nitrogen, such as ammonia, NH_3, that can be used by plants and other biological agents. Insofar as the total amount of N_2 fixed, the biological processes for converting from N_2 to NH_3 are the most significant. In biological nitrogen fixation, microorganisms, either free-living or in symbiosis with plants (mainly in root nodules), reduce N_2 to NH_3 at atmospheric pressure and within the temperature range of 20–37°C. This natural process is to be contrasted with industrial chemical conversion processes which may require up to 300 atmospheres of pressure and a reaction temperature range of 200–300°C.

Biological Nitrogen Fixation. The occurrence and importance to soil fertility of biological nitrogen fixation have been known since the early 1800s. The first major finding did not occur until 1960, however, when it was shown that cell-free extracts of the anaerobic bacterium *Clostridium pasteurianum* could be made to fix nitrogen if molecular oxygen, O_2, were rigorously excluded—and also if pyruvic acid, a source of energy and electrons, was supplied. This finding demonstrated that studies no longer were restricted to whole cells, as previously indicated, but that it should be possible to isolate and chemically identify the components of the nitrogen-fixing system.

The first demonstrable product of cell-free N_2 fixation is NH_3, as had been strongly suggested by previous whole-cell studies. Since the reduction of N_2 to $2NH_3$ requires six electrons and since most electron transfer systems known in biochemical pathways involve either a one- or a two-electron transfer, it could be expected that either six one-electron or three two-electron transfer steps would be involved in nitrogen fixation. This would also suggest the existence of nitrogen compounds of valence states (reduction states) intermediate between N_2 and NH_3. However, no such intermediates have been found even in systems using cell-free extracts.

Because of failure to detect intermediates, attention was focused on the mechanism in extracts of *Clostridium pasteurianum* through which electrons were transferred from pyruvic acid to the nitrogen-fixing system. These investigations led to the discovery and isolation of the new electron carrier ferredoxin (Fd) which functioned by accepting electrons released during pyruvate oxidation by enzymes present in the clostridial extracts. The electrons from reduced Fd were transferred to a variety of different acceptors as directed by the cell. For example, some of the electrons from reduced Fd were transferred to hydrogenase, an enzyme which combined the electrons with protons (H^+) to produce molecular hydrogen, H_2, a major by-product of this anaerobe. Other electrons from reduced Fd were transferred via a flavoprotein carrier to nicotinamide adenine dinucleotide phosphate ($NADP^+$) to yield NADPH, a reduced electron carrier shown to be important in the metabolism of all biological agents. It was also found that electrons from Fd were required for nitrogen fixation when pyruvate was present as supporting substrate.

A major finding was that H_2, through hydrogenase, would act as an electron source for reducing ferredoxin. Thus, in these extracts, H_2 could be used to reduce $NADP^+$ to NADPH and NO_2^- to NH_3, and Fd was necessary as an intermediary electron carrier. Since Fd is required for pyruvate-supported N_2 fixation, it may be expected that H_2 would support nitrogen fixation, since reduced Fd is readily produced from H_2 in these extracts. Molecular H_2 alone, however, did no support N_2 fixation. This suggested either that a component other than reduced Fd was required, or that H_2, although capable of reducing Fd, was inhibitory to N_2 fixation as prior whole-cell studies had indicated. If an additional component were required, it appeared that it was produced from pyruvic acid, since pyruvic acid supported active N_2 fixation.

Several unsuccessful attempts were made to obtain N_2 fixation in extracts to which H_2, N_2, and one of the other products of pyruvate metabolism, ATP, were added. Active N_2 fixation did occur, however, when another product of pyruvate metabolism, acetyl phosphate, was added in addition to H_2 and N_2. When compounds such as ADP were removed from cell extracts by dialysis, no N_2 fixation occurred unless ADP was added together with phosphate, H_2, and N_2. Acetyl phosphate then was acting as a source of ATP. The reason ATP did not work directly was that a continuous supply of ATP was required, and a high concentration of ATP, if added directly to a cell-free extract, was highly inhibitory to N_2 fixation. In whole cells that are fixing N_2, a continuous supply of ATP is made available during sugar metabolism.

Genetic Manipulation. High on the list of many researcher's agendas for projects using the practical application of recombinant DNA research has been the possible development of a living organism that will produce ammonia—in an effort to lessen dependence upon costly and highly energy-consuming synthetic ammonia fertilizers. However, at symposia held on this topic, these achievements are considered by most researchers as quite long-range. There are fundamental problems difficult to overcome, including (1) the possibility that increasing biological nitrogen fixation, for which the plant furnishes the energy, can cause a net decrease in crop yields by depriving the plant of nitrogen for the production of certain critical growth elements; and (2) the very rapid-acting inactivation by oxygen of nitrogen-fixation mechanisms. Cloning techniques may be a path toward introducing nitrogen-fixation genes into certain bacteria. One objective is that of developing new forms of bacteria that will enter into symbiotic relationships with crop plants, such as corn (maize) and wheat, that do not possess their own nitrogen fixation symbionts.

In addition to recombinant DNA and molecular cloning techniques, some scientists have combined their research with more conventional genetic techniques. An *E. coli* plasmid capable of carrying nitrogen-fixation genes of *K. pneumoniae* has been developed. Some researchers also believe that nitrogen-fixation genes may be introduced directly into plant cells to result in a plant that requires no nitrogen fertilizer.

In research activities such as these, much knowledge has been gained concerning the energy needs for biological nitrogen fixation. More energy is used than originally contemplated; for example, 20 moles of adenosine triphosphate (ATP) are required to fix one mole of nitrogen. This contributes largely to the first problem mentioned earlier, namely, the great amount of energy required for the plant to fix its own nitrogen, possibly leading to yield reduction.

The well-known nitrogen fixation by rhizobia (see **Legume**) depends on photosynthesis by the plant. Although the method is essentially impractical, photosynthesis can be increased by blanketing the plant with an atmosphere enriched in carbon dioxide. When this is done in the laboratory, increased legume yields are reported. Some investigators postulate that this effect is the result of a reduction photorespiration, a rather wasteful process in which carbon dioxide gained through photosynthesis is diverted into a series of less productive pathways in the plant. Investigators have also found that 30% of the energy used by the nitrogenase of most rhizobial species goes to producing hydrogen rather than ammonia. Research has also shown that the organisms which perform the nitrogen fixation function in plants are indeed quite diverse in themselves. Thus, new combinations of plants and organisms may increase efficiency in some cases.

As pointed out by Evans-Barber (1977), nitrogen is fixed by a variety of microorganisms in addition to those associated with legumes. Some of these include bacteria located in soils, in decaying wood, and on the surfaces of plant roots. They also include free-living blue-green algae, with fungi, ferns, mosses, liverworts, and higher plants (Hardy-Havelka, 1975). Reviews of numerous nitrogen-fixing organisms are given by Silvester (1976), Dalton (1974), Bond (1974), and Stewart (1974).

Role of Molybdenum in Nitrogen Fixation. Traditionally, molybdenum has been considered a key to the reduction of molecular nitrogen to ammonia by soil- and water-dwelling microorganisms. The metal is believed to be a part of the catalytically active site of nitrogenase, which is the enzyme that accomplishes the reduction. As early as 1980, researchers (North Carolina State University) suggested that the bacterium *Azobacter vinelandii* may have an alternative system for nitrogen fixation, a mechanism that may not require molybdenum. The proposal was regarded with some skepticism until the findings by researchers (Agriculture and Food Research Council Unit of Nitrogen Fixation, University of Sussex, England) in 1985 that confirmed the fact that *A. vinelandii* does have a second fixation system. Mutants of *A. vinelandii* were studied. Genes coding for the nitrogenase proteins were specifically deleted or inactivated. It was found that deletion of all three nitrogenase structural genes did not interfere with the fixation process. However, the process was effective only when molybdenum was *not* present. It was also found that the "wild" type of bacterium must have

Mo in order to reduce nitrogen. Thus, it appears that the alternative mechanism is activated only when the system is subjected to molybdenum starvation. It has been suggested that the alternative system represents an adaptation to molybdenum-poor soils. The extension of these findings to other nitrogen-fixing microorganisms remains to be accomplished. Some further details pertaining to the Sussex investigation are given by Marx (1985).

Madigan (1979) and associates found that photosynthetic purple bacteria can grow with dinitrogen gas as the only source of nitrogen under anaerobic conditions, with light as the energy source. They also found that *Rhodopseudomonas capsula* can fix nitrogen in darkness with alternative energy conversion systems.

See also **Fertilizer.**

Additional Reading

Cheung, H, and J. H. Royal: "Efficiently Produce Ultra-High-Purity Nitrogen On-Site," *Chem. Eng. Progress*, 64 (October 1991).

Graham, P. H., and S. C. Harris: "Biological Nitrogen Fixation," Unipub, New York, 1984.

Metz, C. B.: "Biology of Fertilization, Academic Press, Orlando, Florida, 1985.

Meyers, R. A.: "Handbook of Chemicals Production," McGraw-Hill, New York, 1986.

Staff: "Nitrogen Plant Opens in South Carolina," *Chem. Eng. Progress*, 10 (February 1990).

Staff: "Can Catalytic Combustion in Jet Engines Zap NO_x?" *Chem. Eng. Progress*, 19 (March 1992).

Staff: "Advanced Catalyst Zaps Nitrogen," *Chem. Eng. Progress*, 21 (July 1992).

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

NITROGEN (Ammonia). See **Ammonia.**

NITROGEN (Fertilizer). See **Fertilizer.**

NITROGEN (Fixation). See **Nitrogen.**

NITROGEN GROUP (The). The elements of group 15 of the periodic classification sometimes are referred to as the Nitrogen Group. In order of increasing atomic number, they are nitrogen, phosphorus, arsenic, antimony, and bismuth. The elements of this group are characterized by the presence of five electrons in an outer shell. The similarities of chemical behavior among the elements of this group are less striking than hold for some of the other groups, e.g., the close parallels of the alkali metals or alkaline earths. Although all of the elements of this group have valences in addition to 5+, all do have the 5+ valence in common. Unlike the alkali metals or alkaline earths, for example, the elements of the nitrogen group are not so similar chemically that they comprise a separate group in classical qualitative chemical analysis separations. Three of the five, however, antimony, arsenic, and bismuth are members of the second group in terms of qualitative chemical analysis.

NITROGEN (Photochemical). See **Photochemistry and Photolysis.**

NITROGEN RUNOFF. See **Fertilizer.**

NITROGLYCERIN. See **Explosive.**

NITROPHOSPHATES. See **Fertilizers.**

NMR SPECTROSCOPE. See **Nuclear Magnetic Resonance.**

NOBELIUM. Chemical element symbol No, at. no. 102, at. wt. 254 (mass number of ^{254}No), radioactive metal of the Actinide series, also one of the Transuranium elements. Nobelium has valences of 2^+ and 3^+. In 1957, a group of American, English, and Swedish scientists bombarded a target of several curium isotopes (largely ^{244}Cm) with a beam of ^{13}C ions from the cyclotron at the Nobel Institute for Physics. They obtained a few alpha particles of 8.5 MeV energy and half-life of 10 minutes. This was considered to indicate the presence of element 102 with a probable mass number of 251 or 253. At that time, the element was named nobelium with assignment of the symbol, No. Further experiments at the University of California, however, failed to confirm this discovery. In April 1958, Ghiorso, Sikkeland, Walton, and Seaborg, working with the heavy ion linear accelerator (HILAC) at Berkeley, showed the isotope 102^{254} to be a product of the bombardment of ^{246}Cm with ^{12}C ions. Confirming experiments at Berkeley in 1966 showed the existence of ^{254}No with a 55-second half-life; ^{252}No with a 2.3 second half-life; and ^{257}No with a 23-second half-life. Four other isotopes are now recognized, including ^{255}No with a half-life of 3 minutes.

In 1973, scientists at Oak Ridge National Laboratory and Lawrence Berkeley Laboratory, produced a relatively long-lived isotope of nobelium through the bombardment of ^{248}Cm with ^{18}O ions. A total half-title of 58 ± 5 minutes was computed from the combined data of both laboratories. See also **Chemical Elements.**

Additional Reading

Note: The following *classic references* as listed in prior editions of this encyclopedia are preserved here.

Ditmer, P. F., et al.: "Identification of the Atomic Number of Nobelium by an X-ray Technique," *Phys. Rev. Lett.*, **26**, 17, 1037–1040 (1971).

Fields, P. R., et al.: "Production of the New Element 102," *Phys. Rev.*, **107**, 5, 1460–1462 (1957).

Flerov, G. N., et al.: "Experiments to Produce Element 102," *Sov. Phys. Dokl.*, **3**, 3, 546–548 (1958).

Ghiorso, A., Sikkeland, T., Walton, J. R., and G. T. Seaborg: "Attempts to Confirm the Existence of the 10-minute Isotope of 102," *Phys. Rev. Lett.*, **1**, 1, 17–18 (1958).

Ghiorso, A., Sikkeland, T., Walton, J. R., and G. T. Seaborg: "Element No. 102," *Phys. Rev. Lett.*, **1**, 18–20 (1958).

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

Maly, J., Sikkeland, T., Silva, R., and A. Ghiorso: "Nobelium: Tracer Chemistry of the Divalent and Trivalent Ions," *Science*, **160**, 1114–1115 (1968).

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

Mikheev, V. L., et al.: "Synthesis of Isotopes of Element 102 with Mass Numbers 254, 253, and 252," *Sov. At. Energy*, **22**, 93–100 (1967).

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

Silva, R. J., et al.: "The New Nuclide Nobelium-259" *Nucl. Phys.*, **A216**, 97–108 (1973).

NOBEL PRIZES. In 1895, the will of Alfred Bernhard Nobel, a successful Swedish industrialist and European munitions maker, directed that the great majority of his estate be invested for the purpose of yielding annual prize money to be awarded to persons who, as the result of their personal efforts, made outstanding contributions to the advancement of science, literature, and peace. Initially, the awards were confined to five domains—Physics, Chemistry, Physiology or Medicine, Literature (of an idealistic tendency), and Peace (to promote the fraternity of nations and the abolution or dimunition of standing armies and the formation and increase of peace congresses). In later years, the governors of the fund added Mathematics as a qualifying discipline and, in 1968, a Nobel prize for the Economic Sciences was established. Although a prize for each of the foregoing categories is awarded each year, it is not mandatory that an award be made for each category every year, a factor determined by the governors of the fund. The will became effective when Nobel died on December 10, 1886, but the first prizes were not awarded until 1901 because of the need for legal clarification of Nobel's wishes as demanded by family members who also were mentioned in the will.

In studying the will, attorneys recognized the possible requirement for splitting a given award among two or even three individuals, but with a maximum of three persons per award. Nobel also mentioned certain criteria for selection of award recipients in each field. In connection with the prize for Physics, Nobel mentioned "discovery" or "invention," whereas the words "discovery" or "improvement" were stipulated concerning the prize for Chemistry. In terms of the prize for Physiology and Medicine, the key word was "discovery." These were guiding factors for application by the governors of the fund, especially during the early years.

If adjusted for inflation over the years, the original fund of 27,716,243 Swedish kroners ($7,427,953) was quite significant, especially when allowing for capital gains through investments realized over subsequent years. Money available for the annual prizes is essentially determined by the annual capital generated by the fund, but with the stipulation that at least 10% of that gain be reinvested each year. It is interesting to note that the honor associated with the prize is never split—that is, a Nobel Laureate is so designated even though a prize may be shared by two or three persons. The original will specified that the Royal Academy of Science (Sweden) select winners in Physics and Chemistry, that the Karolinska Institute of Medicine (Stockholm) make the selections in Physiology and Medicine, and that the Swedish Academy (of Letters) select winners in Literature. A committee of the Norwegian Parliament was specified to select winners of the Peace prize, noting "no consideration whatever be paid to the nationality of the candidates."

Over the years, with the addition of new prize categories, the governors of the Nobel Foundation necessarily have amended certain procedures. The Foundation is governed by a five-person board of control made up of one appointment by the King of Sweden and one each as appointed by the aforementioned organizations (now referred to as Nobel Institutes). Members of the selection board are appointed for a period of $4\frac{1}{2}$ years. The science awards are presented on December 10 (anniversary of Nobel's death) of each year at the Stockholm Concert Hall, with personal felicitation of the King of Sweden. The Peace prize is presented in a formal ceremony in Oslo.

The first Nobel prizes were awarded in 1901 to Wilhelm K. Roentgen (discovery of X-rays in 1895), J. H. van't Hoff (chemical thermodynamics and osmotic pressure), and E. A. von Behring (diphtheria antitoxin). Listings of scores of Nobel prizes awarded over a century of progress can be found in a number of references, such as "The Information Please Almanac," 45th Edition, 701–709, Houghton Mifflin Company, Boston, Massachusetts. Specific 1992 winners are described in "U.S. Researchers Gather a Bumper Crop of Laurels," *Science*, 542 (October 23, 1992). An excellent review of the first half-century of the Nobel Foundation is given in an article by George W. Gray, "The Nobel Prizes," *Sci Amer*., 1 (September 1949).

Although a chronological listing of Nobel prizes in the sciences provides a good source of tracing the progress of science over the years, not all major discoveries and inventions have been so honored. There are several outstanding scientists who have not been included. Traditionally, Nobel prizes are given for achievements that date back a few to several years rather than for discoveries of the immediate past.

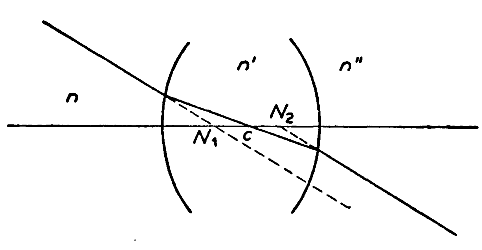
NOBLE FIR. See **Fir Trees.**

NOBLE GASES. See **Inert Gases.**

NOCTILUCA (*Protozoa, Mastigophora*). A genus of protozoans with one phosphorescent species. This minute form is sometimes so abundant in the ocean that the water appears luminous at night.

NODAL LINES. On a vibrating diaphragm, lines along which no vibration takes place. If the diaphragm is circular they consist of two kinds, concentric nodal circles and nodal diameters.

NODAL POINT. Of all the rays that pass through a lens from an off-axis object point to its corresponding image point, there will always be one for which the direction of the ray in the image space is the same as that in the object space (see accompanying diagram). The two points at which these segments, if projected, intersect the axis are called the nodal points, and the transverse planes through them are called the nodal planes. Only if n and n'', the refractive indices in the object and image spaces, are identical are the nodal planes also the principal planes. C is the optical center of the lens.



Nodal points of lens.

NODE. The point, line, or surface in a standing wave system where some characteristic of the wave field has essentially zero amplitude. A node is also a terminal of any branch of a network, or a terminal common to two or more branches of a network. The terms junction point, branch point, and vertex are synonymous.

NODES (Line of). See **Line of Nodes.**

NOISE. Any undesired sound. By extension, noise is any unwanted disturbance within a useful frequency band, such as undesired electric waves in any transmission channel or device. Such disturbances, when produced by other services, are called interference. Noise is also accidental or random fluctuation in electric circuits due to motion of the current carriers. From this concept of noise, the term is used as an adjective to denote unwanted fluctuations in quantities that are desired to remain constant, or to vary in a specified manner. For example, *noise voltage* is a term applied to spontaneous fluctuations in voltage in a component, device or system. The root-mean-square value of these fluctuations then gives a quantitative measure of the noise voltage. With this figure, such other measures as noise power and noise temperature are defined with the necessary stipulation of standard conditions for the acoustic, electric or electroacoustic system.

Background noise is (1) noise due to audible disturbances of periodic and/or random occurrence; (2) in receivers, the noise in absence of signal modulation on the carrier; (3) in recording and reproducing, the total system noise independent of whether or not a signal is present. The signal should not be included as part of the noise.

Thermal noise or Johnson noise is the noise produced by thermal agitation of charges in a conductor. The available thermal noise power produced in a resistance is independent of the resistance value, and is proportional to the absolute temperature and the frequency bandwidth over which the noise is measured, as indicated by the formula

$$N_t = 1.37 \times 10^{-23} T \Delta F$$

in which N_t is the available thermal noise power, T is the temperature of the resistance in degrees Kelvin, and Δf is the bandwidth in cycles per second.

Shot noise is the fluctuation in the current of charge carriers passing through a surface at statistically independent times. It has a uniform spectral density W_i given by

$$W_i = \frac{eI_o}{2\pi}$$

Random noise, exemplified by thermal noise and shot noise, has a uniform energy versus frequency distribution. *White noise* (or *Gaussian noise*) is a random noise having a constant energy per unit bandwidth that is independent of the central frequency of the band.

Over a number of years, society has been plague with ever-increasing noise from traffic, machinery, overamplified and discordant music, rocket and jet engine blasts, and sonic booms. Excessive and constant exposure of the ear to noise can and does cause impairment or loss of hearing. In addition other physical and frequently psychological changes can also occur as the result of noise disturbing sleep, impairing efficiency, and otherwise causing emotional disturbances. During wartime, military personnel often receive partial or total hearing loss after exposure to blasts and gunfire. Nearly 60,000 veterans of World War II, for example, have received compensation for ear damage considered to be a service-connected disability. Occupational deafness (earlier referred to as boiler maker's deafness) is a hearing loss resulting from prolonged exposure to industrial noise. Such damage is permanent, and may be partial or total. Overamplified music, particularly favored by teenagers also is a cause of hearing impairment although the problem has not been subject to the same degree of analysis as in the case of industrial noises. See also **Hearing and the Ear; Supersonic Aerodynamics.**

NOISE LEVELS FOR VARIOUS SOURCES AND LOCATIONS

Description of Noise	Noise Level (dB)
Threshold of hearing	0
Rustle of leaves in gentle breeze	10
Quiet whisper (distance of 5 feet)	10
Average whisper (distance of 4 feet)	20
House in country (average situation)	30
House in city (average situation)	40
Apartment (average situation)	40
Hotel	42
Theater (between performances)	42
Small retail establishment	52
Commercial garage	55
Medium-size office	58
Residential street	58
Restaurant	60+
Medium-size retail establishment	62
Factory or warehouse office	63
Large retail establishment	63
Ordinary conversation (distance of 3 feet)	65
Large office	65
Traffic on busy street	68
Factory (light-to-medium work)	78
Riveter (distance of 35 feet)	97
Hammer blows on steel plate (distance of 2 feet)	114
Threshold of pain	130

One of the frustrating results of noise is the masking effect it produces in reducing the intelligibility of speech. For example, if the speaker and listener are separated by 5 feet (1.5 meters), the levels of noise that will barely permit reliable word intelligibility are 50 decibels (dB) for normal conversation; 57 dB for raised speech; 63 dB for very loud speech; and 69 dB for shouting. As shown by the accompanying table, these levels are approached or exceeded in several day-to-day industrial and commercial activities. See also **Acoustics.**

Additional Reading

Cunniff, P. F.: "Environmental Noise Polution," Wiley, New York, 1977.

Fader, B.: "Industrial Noise Control," Wiley, New York, 1981.

Kryter, K. D.: "The Effects of Noise on Man," 2nd Edition, Academic Press, Orlando, Florida, 1985.

Lord, H., and H. A. Evensen: "Noise Control for Engineers," McGraw-Hill, New York, 1980.

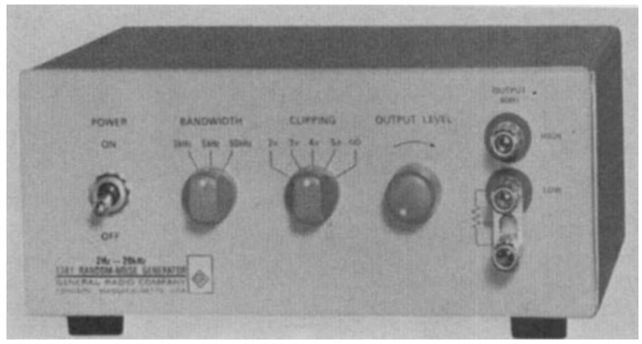
Ott, H. W.: "Noise Reduction Techniques in Electronic Systems," Wiley, New York, 1976.

White, R. G., and J. B. Large: "Noise and Vibration," Wiley, New York, 1983.

NOISE GENERATOR. Electronic noise, by definition, is an unwanted disturbance and its reduction in communications circuits is a constant aim of the electronics engineer. When supplied by a properly controlled generator, however, noise becomes a very useful test signal. Random noise contains no periodic components and its future value is completely unpredictable. It is described by its amplitude distribution and its spectrum. In its most common form, it has a gaussian distribution of amplitudes. Two spectral varieties are useful: (1) "white noise" has a uniform spectral level over a specified frequency range, i.e., equal energy in each hertz of frequency; (2) "pink noise" has equal energy in each octave and, therefore, the energy in each hertz decreases at a rate of 3 decibels per octave. The random-noise signal, embracing a wide range of frequencies and having a randomly varying instantaneous amplitude, closely approximates the signals normally encountered in many electronic circuits and, particularly, in busy communication systems.

Specific applications for electronic noise generators include, as a broadband signal source: (1) intermodulation and cross-talk tests, (2) simulation of telephone line noise, (3) measurements on servo amplifiers, (4) noise interference tests on radar, (5) determining meter response characteristics, (6) setting transmission levels in communication circuits, (7) frequency response measurements; and, as a signal source, for (8) the measurement of reverberation, (9) sound attenuation of ducts, walls, panels, or floors, (10) acoustical properties of materials, (11) room acoustics, and (12) classroom or laboratory demonstrations. With a suitable power amplifier, such devices (13) can be used to drive a loudspeaker to produce high level acoustic noise for fatigue testing of structures and components, and (14) can drive a vibration shaker for structural tests of components and assemblies.

Two noise sources for the audio-frequency range are in common use. One is the semiconductor diode, operated at low current in the reverse breakdown mode. In the instrument illustrated, noise from a semiconductor diode is amplified, filtered to establish the spectral shape, and further amplified to produce an output level of 3 volts. Internal circuits provide clipping of the peaks of the noise, if desired, or change the spectrum from white to pink noise as required.



Random-noise generator. (*GenRad, Inc.*)

The other noise source is the pseudo-random noise generator, constructed of digital circuit elements connected as a shift register with feedback connections. When the feedback and initialization are such as to produce the maximum length sequence when the shift register is clocked, the rectangular wave output can be taken from the shift register and filtered with a low-pass filter to produce a waveform with good approximation to Gaussian or normal amplitude distribution. Such noise is called pseudo-random noise, because the pattern is precisely repeated each time the shift register sequence is repeated.

The most common noise source for the radio-frequency range is the temperature-limited thermionic diode. Its shot noise, although very low in level, is used because of its excellent spectral flatness and true Gaussian amplitude distribution.

James J. Faran, Jr. GenRad, Inc., Concord, Massachusetts.

NOISE (Statistics). Disturbance terms, analogous to noise in the engineering sense, which appear in time-series. They are usually regarded as random elements superposed on the systematic components of the series and have a tendency to obscure its nature.

NOISE (Supersonic). See **Supersonic Aerodynamics.**

NOMENCLATURE (Organic Chemistry). See **Organic Chemistry.**

NONDESTRUCTIVE TESTING (NDT). The examination of materials and objects for the purpose of detecting defects without in any way harming the test object. NDT contrasts vividly with destructive testing methods, which chemically consume or physically damage the test object, rendering it unfit for use. Whereas destructive testing must be confined to statistical sampling procedures, NTD enables 100% on-line inspection if desired. The trend in recent years has been in this direction, with emphasis on automating and increasing the speed of NDT operations. Another significant trend has been that of *testing work in progress*, as contrasted with earlier procedures which concentrated on testing raw materials and final products. In this way, very helpful information for step-by-step quality control can be provided. There remain, of course, numerous examples of where statistical destructive testing is needed—for example, in determining the ultimate compressive and tensile strengths of materials and parts or checking the corrosion resistance of materials. In recent years, it has proven possible to combine the results of NDT with computerized simulation in some instances to predict failure of test objects under certain conditions. For obvious economic reasons, NDT is preferred by manufacturers over destructive testing this accounts for high acceptance and many advancements which have occurred in NDT methods. For research and development applications, nondestructive methods are sometimes referred to as NDE (nondestructive evaluation).

Traditionally, NDT has been associated with metals and materials of construction for finding potentially unsafe conditions, such as cracks, voids, holes, inclusions, and other inconsistences, as may be found in metal sheets, plates, bars, tubes, castings, forgings, and weldments. Such defects may arise from faulty manufacturing, or from later use, as the result of corrosion, abrasion, vibration, mishandling, and inattention to required maintenance procedures. In recent years, the applications for NDT have broadened to include all manner of materials—films, coatings, polymers, composites, and ceramics as encountered in a wide variety of industries, including numerous uses in the electronics manufacturing industry. Also, NDT is widely used for on-site inspection of large and heavy equipment, which cannot be detached for testing, after installation, but where periodic checks are required. Examples include the inspection of weldments in pipelines, aircraft engines and structural components, military equipment, bridge structures, etc.

During the last half of the 1980s and carrying well into the 1990s, NTD has enjoyed the benefits of measurement and computer technologies that have contributed immensely to the speed, accuracy, and reliability of NTD, even though instrument costs have risen markedly as a result. However, the ability to make more measurements within shorter periods of time probably has not increased the unit costs proportionately. A number of measurement techniques entirely new to the NDT field have been added to increase the variety of choices.

Radiographic Methods

Radiographic (x- and gamma-ray technology) method using film was one of the earliest NTD schemes used. Although early systems are undergoing modernization, this comparatively simple method still enjoys acceptance for certain applications. This basic technique has taken on a number of new formats.

Film Images. Images made by the traditional film technique are shown in Fig. 1. To reduce costs and meet environmental restrictions, a dry-silver system was introduced in 1991. The system produces a silver-based image without the use of wet chemistry, using photothermographic technology. The image is developed on exposed film by thermal energy rather than by the traditional method of immersing film in a liquid developer and fixer. Three elements required for dry processing are a specially coated film, fluorescent exposing screen, and a thermal processor. The film has a translucent polyester base similar to that of conventional film. Its ultrafine grain produces detailed images of archival quality.

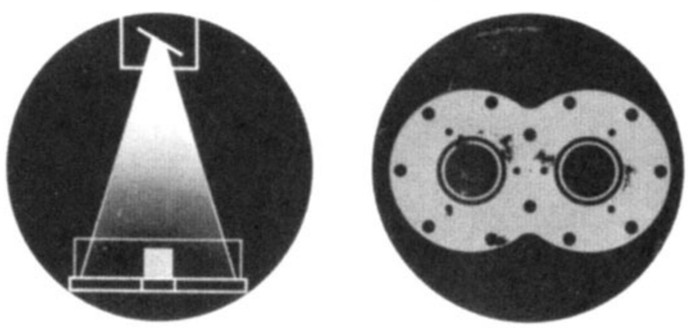


Fig. 1. X-rays or gamma rays are used to create a shadow image of light and dark that reveals any flaws or inclusions in a test part.

Traditional radiographic methods use two-dimensional film to record the attenuation of X-rays passing through a three-dimensional object. The result is a shadowgraph in which all object features are superposed. To improve the totality of information obtained, backscattering methods were introduced several years ago.

Principal applications for radiography include the inspection of castings, electrical assemblies, weldments, small, thin, and complex wrought products, some nonmetallics, solid propellant rocket motors, cans or containers, composites, and nuclear reactor fuel rods, among many others.

Chronology of Radiographic Methods. In 1985, researchers at John Hopkins University described a *flash X-ray system* that uses increased-power X-ray sources to generate very intense short pulses. High-gain X-ray intensifier detectors are used. Exposure times as short as 30 ns are possible and thus microstructural changes due to explosions, heat pulses, and shock waves can be detected. An indirect and direct method are used. In the indirect method, the X-ray diffraction image is converted into a visible light image by a fluorescent screen. The researchers have found that for the indirect method, a multiple-stage image-intensifier system coupled to an external fluorescent screen is the most sensitive and almost instantaneous system. Multiple stages of amplification allow individual X-ray photons to be detected. In more advanced systems, there is inclusion of a microchannel plate where electrons strike the output phosphor and are converted into a strong, visible image.

In the direct method, an X-ray-sensitive vidicon TV camera directly converts the X-ray image into an electronic charge pattern on a photoconductive target, which is read out by a scanning electron beam and displayed visually on a TV monitor.

In addition to testing uses per se, flash-X-ray techniques have been used to study the orientation of single crystals, to study lattice rotation accompanying plastic deformation, to measure the grain boundary migration during recrystallization annealing, and to determine the physical state of exploding materials.

In another technique known as X-ray transmission asymmetric crystal topography, changes in defect structure during polymerization of single crystals have been studied.

Digital Radiography. In this technique, the traditional film is replaced by a linear array of detectors and the X-ray beam is collimated into a fan beam. The object is moved perpendicularly to the detector array, and the attenuated radiation is sampled digitally by the detectors. Data are processed by stored information in the computer's memory to yield a two-dimensional image of the part being inspected.

X-Ray Computed Tomography (CT). In computed tomography, penetrating radiation from many angles is used to reconstruct cross-sectional images of an object. The advantages of CT are exemplified by the inspection of aircraft/aerospace castings for internal defects. Advantages of CT include increased reliability, elimination of unnecessary rejects, and wider use of castings instead of forgings and parts machined from wrought stock. CT has been found to have greater sensitivity (dependent on part size and geometry) than conventional film. CT can spatially define flaw distribution. Aerospace test engineers claim that castings can be measured with an accuracy of better than 0.05 mm (0.002 in), but is adversely affected by the amount of image noise and the edge-detection method used. Computed tomography systems are costly. The general principles of CT are described in article on **X-Ray Scan and Other Medical Imagery**.

Ultrasonic Methods

Typically, ultrasonic images are produced by mechanically scanning an ultrasonic transducer in a raster pattern over an area of a structure and then displaying the reflected or transmitted energy in a suitable format. Usually, the scan is performed in a tank of water or with some form of squirter nozzle. The liquid medium serves to transmit the ultrasonic energy from the transducer into the test material. Conventionally, the data are displayed as C-scans (a plan view image where a color scale is used to display signal amplitude or depth information) or as B-scans (image of a cross section at one particular location of interest, typically with a color indicating signal amplitude).

As indicated by Fig. 2, there are several testing modes: (1) pulse-echo mode, (2) through-transmission mode, (3) reflector-plate mode, and (4) angle-beam mode.

Sonic (<0.1 MHz) and ultrasonic (0.1 to 25 MHz) radiation have been used for many years in NDT. In a simple testing scheme, sonic or

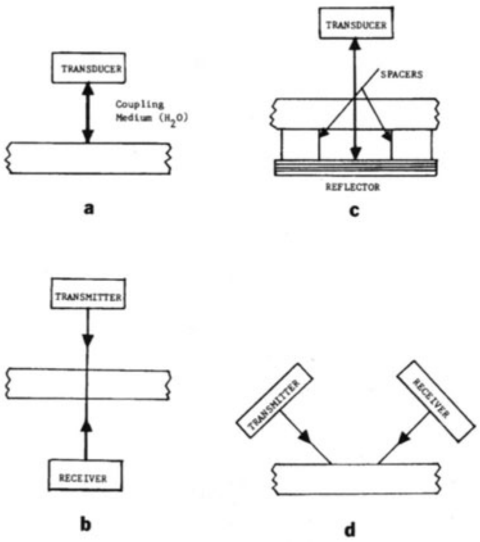


Fig. 2. Ultrasonic NDT methods: (a) pulse-echo; (b) through-transmission; (c) reflector-plate (double-through transmission); and (d) angle-beam.

ultrasonic vibrations are generated and sent by way of a pulse beam through the part to be tested. The beam travels unimpeded through large parts, may be angled for testing sheet stock, and can impact materials immersed in a liquid. Any flaw reflects vibrations back to the instrument, which indicates the location and size of the discontinuity on a CRT (cathode-ray tube). Access is required to only one side of the material being tested. Although energy can be lost from the ultrasonic beam due to geometrical effects, these can be controlled to increase the sensitivity of attenuation measurements. Hence, microstructural alterations, such as microcracks, foreign particles, precipitates, grain boundaries, interphase boundaries, and dislocation defects, can be detected. Research has shown that attenuation measurements have detected microstructural change during fatigue testing, therefore giving early warning to fatigue-induced failure, as well as measuring oxygen content in titanium welds.

Acoustic methods are applicable to numerous kinds of materials. The method can be used, for example, to reveal fiber/matrix bond strength in polymer-matrix composites. In one method (Wan-li Wu, National Institute of Standards and Technology), a continuous wave argon-ion laser is used to heat a very small area of the composite. The resulting thermal expansion between fiber and resin produces a measurable change due to debinding. Conventional methods of evaluating bond strength are time consuming and tedious. Instead of measuring the thermal stress, the laser power level at which debonding occurs is used as the index of debonding stress. Although sonic scanning techniques can be used to detect voids and cracks at interfaces in polymer-matrix composites, they do not measure the strength of interfacial bonds.

As pointed out by D. Sturges (General Electric Aircraft), "Modern ceramic materials offer many attractive physical and mechanical properties for use in a rapidly growing variety of industrial applications. The critical nature of many applications, however, imposes technical challenges in manufacturing and inspection. One nondestructive evaluation (NDE) technique of major relevance to inspecting ceramics is ultrasonic microscopy (also termed acoustic microscopy), that is, the use of tightly-focused, high-frequency sound beams to form images of the point-to-point reaction of a material to periodic stress waves. This technique offers high sensitivity for the detection of small defects, and often is a complementary technique to X-ray inspection."

Computer-assisted ultrasonic microscopy (CAUM) has been of particular significance in the testing of new materials developed for more fuel-efficient engines, wherein one objective is that of maximizing the high thermal efficiency of gas turbine engines by way of incorporating high-temperature ceramic components and exhaust-heat recovery. The object of NDE is that of assuring that ceramic components are free of both surface and internal flows that limit component life. Surface flaws can be generated during production by machining and normal handling.

Penetrant Method. This method does not depend upon radiation interactions with the test object and is essentially noninstrumental. A special penetrant substance is applied freely on the test object and allowed to work into tight cracks. See Fig. 3. The penetrant is removed from all surface areas and the piece is sprayed with a developer. The developer dries to an even white coating, while the penetrant bleeds up from any flaws through the developer, forming bright-red or fluorescent indications on the white surface. The size of the defect is indicated by the richness of color, speed of bleed-out, and dimensions observed.

Because of environmental concerns, a new generation of biodegradable penetrants having sensitivity levels ranging from 1 to 4 have been developed. The new penetrants are water washable and, in most instances, can be directly discharged into sewers. They are free of petroleum-based solutions.

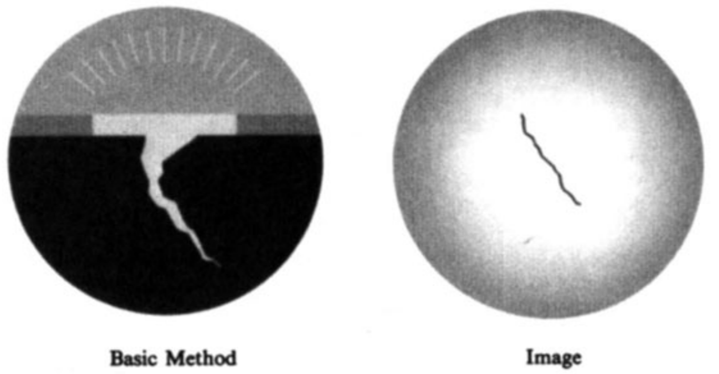


Fig. 3. Penetrant method for detecting flaws.

Magnetic-Particle Method. This method makes use of iron powder to reveal the leakage magnetic field created at a flaw or break when any part is magnetized. The familiar horseshoe magnet best illustrates this principle. (1) If a horseshoe magnet is bent into a circle, the field between the ends attracts and holds magnetic iron powder. (2) If a magnet is made completely closed, the field will be contained entirely within the ring and no iron powder will be attracted. (3) However, if the round magnet is cracked, poles are created at the break, and iron powder is instantly attracted to the cracked area to pinpoint the defect. See Fig. 4.

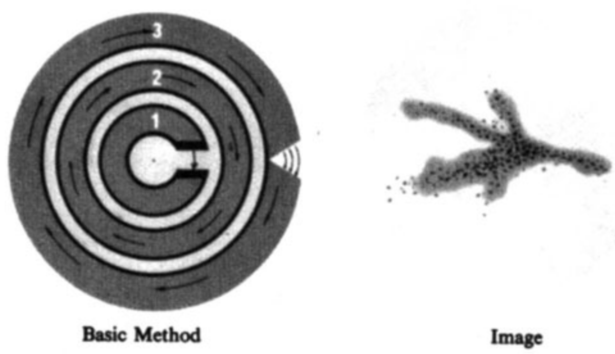


Fig. 4. Magnetic particle method for detecting flaws.

Eddy-Current Methods. This is one of the earliest NDT methods and is still used. Basically, this method reveals any differences in electrical impedance between parts to be tested and a reference sample. Parts to be examined are passed through a coil or explored with a probe, and a trace appears on a CRT. Since magnetic and electrical characteristics are closely related to metallurgical quantities, a trace position or pattern or a meter reading clearly shows variations in metal hardness and composition, as well as defects. Both ferrous and nonferrous parts can be tested, and various coils, probes, and detector tips are available.

Aside from more sophisticated electronics, a major contribution to improve eddy-current instrumentation has come from the development of the eddy current resonance digitizing (ECRD) method. With this method, eddy-current instrumentation can separate nonferrous alloys based upon characteristics other than simply their conductivity. See Fig. 5.

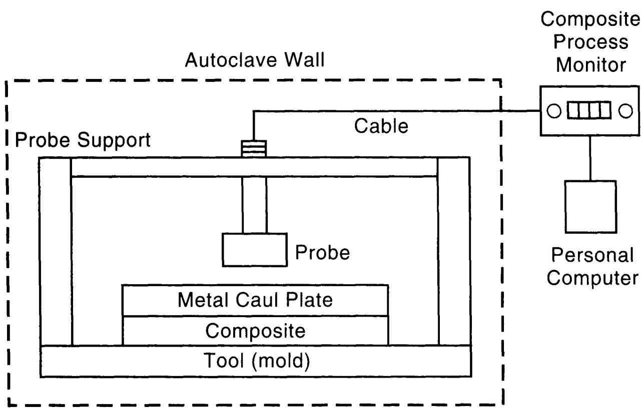


Fig. 5. Use of eddy-current testing for monitoring a composite cure. Typically, eddy-current testing uses an electronic instrument having a small probe on the end of a flexible electrical cable. The probe is placed either against or close to the target. The target must be electrically conductive to allow the generation of eddy currents. The time-variant nature of the probe's magnetic field causes electric currents to flow in the target material. Higher field frequencies or a more electrically conductive material increase the depth of the eddy-current penetration into the material. The concentration of eddy currents near the surface of the material is referred to as the "skin effect." Eddy currents generate their own magnetic field that opposes the probe's magnetic field. Detection circuits in the instrument sense the impedance changes in terms of phase/amplitude changes in probe-coil voltage. (*Suggested by Bar-Cohen and Nguyen.*)

NDT Outlook

Improvements in current, established technologies and the introduction of new ways to test materials, nondestructively are expected to continue apace during the remaining years of the 1990s. One promising method is *positron annihilation*. The positron is the antiparticle of the electron; thus a positron/electron pair is unstable and will annihilate. In this process, two gamma rays at approximately 180° to one another are emitted from the center of the mass of the pair. A very slight departure from 180° is directly proportional to the transverse component of the momentum of the pair. The momenta of the electrons involved in such collisons can be calculated from the geometry and intensity of the gamma rays. The dynamics of the electron/positron system underlie the use of the technique for the study of defects in materials.

Laser-based *shearography* is also in early stages of development.

Additional Reading

Adams, T. E., and A. C. Wey: "Nondestructive Sectioning: Alternative to Physical Sectioning," *Advanced Materials & Processes*, 54 (February 1992).
Akuezue, H. C., and S. K. Verma: "Positron Annihilation: NDE at the Atomic Level," *Advanced Materials & Processes*, 26 (March 1992).
Altshuler, T. L.: "Atomic-Scale Materials Characterization," *Advanced Materials & Processes*, 18 (September 1991).
Bar-Dohen, Nguyen, K. H., and R. Botsco: "Eddy Currents Monitor Composites Cure," *Advanced Materials & Processes*, 41 (April 1991).
Bindell, J. B.: "Elements of Scanning Electron Microscopy," *Advanced Materials & Processes*, 20 (March 1993).
Carter, G. F., and D. E. Paul: "Materials Science and Engineering," ASM International, Materials Park, Ohio, 1991.
Cormia, R. D.: "Problem-Solving Surface Analysis Techniques," *Advanced Materials & Processes*, 16 (December 1992).
Dulski, T. R.: "Residual-Element Analysis: Measuring the Minuscule," *Advanced Materials & Processes*, 20 (February 1992).
Evans, N. J.: "Impedance Spectroscopy Reveals Materials Characteristics," *Advanced Materials & Processes*, 41 (November 1991).
Michaels, T. E., and B. D. Davidson: "Ultrasonic Inspection Detects Hidden Damage in Composites," *Advanced Materials & Processes*, 34 (March 1993).
Place, T. A., et al., Editors: "Computer-Aided Microscopy and Metallography," ASM International, Materials Park, Ohio, 1990.
Prask, H. J.: "Neutron Probes Tackle Industrial Problems," *Advanced Materials & Processes*, 26 (September 1991).
Staff: "Nondestructive Evaluation and Quality Control," ASM International, Materials Park, Ohio, 1988.
Staff: "Testing for Materials Selection," *Advanced Materials & Processes*, 5 (June 1990).
Staff: "Computed Tomography Details Casting Defects," *Advanced Materials & Processes*, 54 (November 1990).
Sturges, D.: "Sounding Out Ceramic Quality," *Advanced Materials & Processes*, 35 (April 1991).
Staff: "Nondestructive Examination," *Advanced Materials & Processes*, 63 (January 1992).
Webb, S. C.: "PCs Help Optimize Materials Testing," *Advanced Materials & Processes*, 21 (November 1991).
Wu, Wen-li: "Acoustic Emissions Reveal Fiber/Matrix Bond Strength," *Advanced Materials and Processes*, 39 (August 1991).

NONEUCLIDEAN GEOMETRY. See **Geometry.**

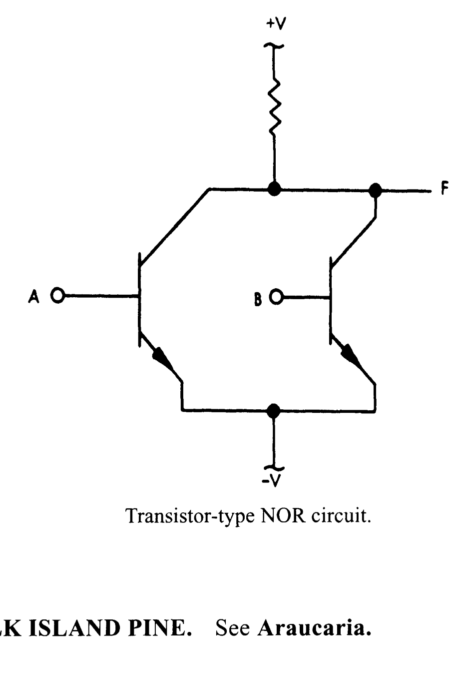
NONNEWTONIAN LIQUIDS. See **Viscosity.**

NONNUTRITIVE SWEETENERS. See **Sweeteners.**

NONSINUSOIDAL WAVE. Any periodic wave which is not a pure sine wave. Such waves may, however, be analyzed into numerous sine components, utilizing Fourier series, and often circuits may be analyzed by considering these one at a time.

NOR (Circuit). A computer logical decision element which provides a binary 1 output if all the input signals are a binary 0. This is the overall NOT of the logical OR operation. Output F is positive only when both transistors are cut off. This occurs when both inputs A and B are negative. See accompanying figure.

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



Transistor-type NOR circuit.

NORFOLK ISLAND PINE. See **Araucaria.**

NORMAL CONCENTRATION. A one normal solution (often abbreviated 1*N*) contains one gram-equivalent weight of a particular substance dissolved in 1 liter of *solution*. The equivalent weight of a substance may be defined as that weight of the substance which will involve, in a chemical reaction, one atomic weight of hydrogen, or that weight of any other element or portion of a substance which, in turn, would involve in reaction one atomic weight of hydrogen.

As an example, the chlorine atom of potassium chloride (KCl) also is found in hydrochloric acid (HCl) in combination with one hydrogen atom. Thus, the gram-equivalent weight of KCl is 74.555, which is the same as its gram-molecular weight. A one normal solution of KCl will contain 74.555 grams of the salt per liter of solution.

For a particular solution, the molar and normal concentration are the same only when the gram-molecular and gram-equivalent weights are the same. Sulfuric acid H_2SO_4 represents a case where these values are not the same. This acid contains two active hydrogen ions and, therefore, its gram-equivalent weight is one-half of its gram-molecular weight. Phosphoric acid H_3PO_4 contains three active hydrogen ions. Consequently, the gram-equivalent weight for this acid is one-third that of the gram-molecular weight. Calcium hydroxide $Ca(OH)_2$ contains two active hydroxyl ions, each being equivalent to a hydrogen ion. Therefore, the gram-equivalent weight of $Ca(OH)_2$ is one-half of its gram-molecular weight.

NORMAL EQUIVALENT DEVIATE. The normal equivalent deviate of a proportion p is the deviate in a normal distribution with unit variance that exceeds a proportion p of the total frequency. Thus N.E.D. $(p) = x$ where

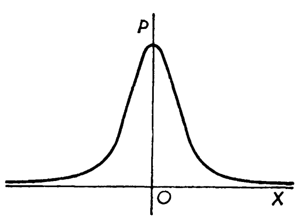
$$p = (2\pi)^{-1/2} \int_{-\infty}^{x} \exp(-\tfrac{1}{2}t^2)\, dt$$

For ease in computation, x is often replaced by y, the probit of p, where $y = x + 5$.

NORMAL (Gaussian) DISTRIBUTION. One of the standard distributions of statistical theory. The mathematical form may be written

$$f(x) = \frac{1}{\sigma(2\pi)} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

This has mean μ and variance σ^2 and the form of the frequency curve is shown in the accompanying figure.



Typical Gaussian distribution curve.

The distribution was considered by Laplace and Gauss and is known by various names, that of "normal" being conferred by Karl Pearson, although he admitted that the practical occurrence of data of exactly normal type was somewhat abnormal. However, a number of distributions occurring in practice are approximately of the normal form, such as that of heights of men or women, or errors of measurement in reported observations of a magnitude.

The distribution owes its theoretical importance to a number of features, notably that under the **Central Limit Theorem** (q.v.) a great many of the statistics in common use tend to have a normal distribution in large samples, so that asymptotically the standard error can be used to set probability limits to estimates of parent parameters even in data from non-normal populations.

The distribution is a simple transfer of the "error curve." See **Error.**

The distribution can be generalized to the case of p variables, being then of the form

$$f \infty \exp\left\{-\frac{1}{2}\sum_{j,k=1}^{p} \alpha_{jk}(x_j-\mu_j)(x_k-\mu_k)\right\}, \quad -\infty \le x_j \le \alpha$$

That is to say, the quantity in the exponential is a general quadratic form, limited only by the fact that the form must be positive definite in order to ensure the convergence to unity of the total frequency.

Sir Maurice Kendall, International Statistical Institute, London.

NORMAL (Geometry). Perpendicular. At right angles to a given line or plane.

NORMALIZE (Mathematics). 1. To change in scale so that the sum of squares, or the integral of the squares of the transformed quantity is unit. (See **Orthogonal Function.**) 2. To transform a random variable so that the resulting random variable has a normal distribution. 3. In computer operations, to adjust the exponent and coefficient of a floating-point result so that the coefficient is in the prescribed normal range. Also called standardize.

NORMALIZING. See **Iron Metals, Alloys, and Steels.**

NORMAL (Principal, to a Curve at a Point *P*). The normal to the curve at *P* which lies in the osculating plane at *P*. A unit vector in the direction of the principal normal is called the *unit* (*principal*) *normal*. It is usually taken as directed from *P* to the concave side of the curve.

NORTH. For many centuries, north has been the fundamental direction used by navigators and surveyors. During this long period, a loose usage of the term has become prevalent. It seems desirable to set down certain standard meanings accepted by the majority of modern navigators and astronomers.

True north (unless a qualifying adjective is used with north, true north is to be assumed) is the direction along the geographical meridian of the observer, in the plane of the observer's horizon, toward the north pole of rotation of the earth. When the observer is facing the setting sun, the north pole is to his right. *Compass north* is the direction of the plane of the horizon toward which the north-seeking end of the compass points. Unless otherwise stated, compass north refers to north defined by the magnetic compass; if another type of compass is used, it should be clearly indicated, e.g., gyrocompass north, etc. *Magnetic north* is the direction in the plane of the observer's horizon toward the north magnetic pole of terrestrial magentism. For methods of conversion from any one of these three "norths" to any other, see **Compass (Navigation).** See also **Navigation.**

NORTH AMERICAN HIGH. See **Atmosphere (Earth).**

NORTH ATLANTIC FLOUNDER. See **Flatfishes.**

NORTH ATLANTIC WATERS. In studies of oceanography, numerous bodies of ocean water are designated. The principal bodies in the North Atlantic are:

North Atlantic Central Water. A shallow oceanic water mass extending roughly from the southern parts of Greenland and Iceland to a region described by a line drawn from the northern end of South America to Africa. Temperature range is 8–19°C (46.4–66.2°F), salinity ranges from 35.1 to 36.7%.

North Atlantic Deep and Bottom Water. This dense ocean current arises in the Atlantic Ocean near the southeastern tip of Greenland where it meets the warmer water of the Gulf Stream below which it sinks to a depth of from 7,000 to 13,000 feet (2,100 to 3,900 meters) as it creeps southward. This water has been traced as far south as 60 degrees where the colder and heavier Antarctic water forces it to the surface.

North Atlantic Intermediate Water. An oceanic water mass lying at depths between the North Atlantic Deep and bottom water and the North Atlantic central water. Temperature range is from 2.5 to 4.0°C

(36.5 to 39.2°F), salinity range is from 34.7 to 34.9%. The area of this water is more limited than those of the water masses above and below it. It has a high oxygen content.

NORTHERN LIGHTS. See **Aurora and Airglow.**

NORTHERN PIKE. See **Pike.**

NORTH PACIFIC WATERS. In studies of oceanography, numerous bodies of ocean water are designated. The principal bodies in the North Pacific are:

North Pacific Central Water. Due to the great size of the Pacific Ocean, it contains more well-developed oceanic water masses than the Atlantic Ocean. Thus, there are both eastern and western north Pacific central surface water masses. They are relatively shallow, extending from the subarctic Pacific water on the north to the Pacific equatorial water on the south, and covering the width of the ocean except for a transition zone on the eastern side. Their temperature ranges are 8 to 18°C (46.4 to 64.4°F) and their salinity ranges from 33.8 to 34.9%, the lower values of each existing at lower depths (400 to 700 meters) (1,320 to 2,310 feet) where they meet the north Pacific intermediate water.

North Pacific Deep Water. Measurements at great depths show the existence of an oceanic water mass of practically constant salinity and low temperature below 2,500 to 3,000 meters (8,250 to 9,900 feet) in the south Pacific Ocean, and still deeper in the north Pacific Ocean. Since the Bering Straits are so narrow and shallow, this mass cannot be produced by currents from the Arctic Ocean, as in the north Atlantic deep and bottom water. Since the oxygen content of the Pacific deep water is less in the north Pacific than the south Pacific Ocean, this water mass is believed to be supplied from the antarctic bottom water or the Atlantic deep and bottom water.

North Pacific Intermediate Water. An oceanic water mass lying below the north Pacific central and equatorial waters at depths ranging from 600 to 800 meters (1,980 to 2,640 feet) in the north and to about 200 meters (660 feet) and 1,000 meters (3,300 feet) (two layers) in the south. It is characterized by low salinity and especially low oxygen content.

NORWALK VIRUS. An epidemic of winter vomiting disease in 1968 involved many residents of Norwalk, Ohio. At that time, the causative agent was unknown. However, the illness resembled a vomiting and diarrhea syndrome observed in some regions as early as 1929. The syndrome occurred during the winter and affected mainly teenagers and adults. Similar episodes among close-living groups of people were reported from boarding schools in England during the 1950s. A number of enteroviruses and rotaviruses were identified in earlier years, such as the coxsackieviruses and echoviruses, among others, but not all epidemics had been satisfactorily explained. The Norwalk agent did not exactly fit any of the prior patterns.

Field investigators from the Center for Disease Control (Atlanta, Georgia) were despatched to Norwalk to collect stool specimens and early and convalescent blood. These were frozen for future examination. The National Institute of Health (United States) established a program involving volunteers and electron microscopic examinations, directed at identifying the Norwalk agent. A specific virus was identified. It resembles paroviruses, measures about 27 nanometers (0.027 micrometer) and has since been officially called the Norwalk agent. Three serologically different forms have been identified. One of these is now called the Hawaiian agent. More recently, additional varying agents have been identified in England. A widespread outbreak of Norwalk virus gastroenteritis in Australia was attributed to contaminated shellfish.

Persons infected with the Norwalk agent display (after 12 hours incubation) fever, vomiting, diarrhea, cramps, malaise, and leukopenia over a period of 2 to 3 days. Biopsy of the intestine during the acute phase of the disease has demonstrated that the villi of jejunum become flattened and infiltrated with mononuclear cells. Absorption of xylose and fat is reduced during and after the illness. Treatment is supportive with rehydration therapy for any major loss of fluids. Experience with volunteers indicates that natural immunity as the result of infection may persist only for a matter of weeks or months in many cases. Several volunteers who were rechallenged after 27 and 42 months developed symptoms of the disease. Because of wide variation in the results of the volunteer program, some authorities believe that there may be genetic differences in susceptibility.

Additional Reading

Blacklow, N. R., Dolin, R., and D. S. Fedson: "Acute Infectious Nonbacterial Gastroenteritis: Etiology and Pathogenesis," *Ann. Intern. Med.*, **76**, 993 (1972).
Melnick, J. L.: "Enteroviruses. Viral Infections of Humans: Epidemiology and Control," (A. S. Evans, editor), Plenum Medical Book Co., New York, 1976.

R. C. V.

NORWAY SPRUCE. See **Spruce Trees.**

NORWEGIAN TOPKNOT. See **Flatfishes.**

NOSE FLY (*Insecta, Diptera*). Also known as the *sheep bot*, this species, *Oestrus ovis* (Linne), acts against sheep, goat, and wild deer and is found throughout North America; it is particularly abundant in Idaho, Montana, New Mexico, and Texas, and generally in the areas lying between these states. The nose fly strikes at the nose of the animal and deposits eggs in the nostrils. The maggots lodged in the nostrils and head sinuses create an inflamed condition and accompanying catarrhal discharge. The insect is known, on occasion, to attack humans.

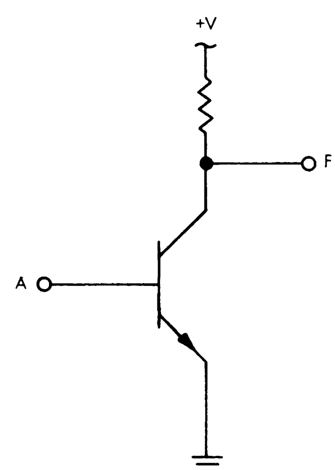
Where the insect is present, the nostrils of the animal should be treated with pine tar. An early application is required in mid-April in most areas. Specially-treated salt logs are obtainable which permit the animals to smear their noses with pine tar in an "automatic" fashion. In times of abundant presence of the flies, some producers herd the animals into darkened retreats where the flies are not active.

NOSE (Odor Receptor). See **Flavorings; Olfactory Organ.**

NOT (Circuit). Also known as an inverter circuit, this is a circuit which provides a logical NOT of the input signal. If the input signal is a binary 1, the output is a binary 0. If the input signal is in the 0 state, the output is in the 1 state. Referring to the accompanying figure, if A is positive, the output F is at 0 V inasmuch as the transistor is biased into conduction. If A is at 0 V, the output is at +V because the transistor is cut off. Expressed in Boolean algebra, F = A′, where the prime denote the NOT function.

See also **Diode Transistor Logic.**

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



Inverter or NOT circuit.

NOTOCHORD. A longitudinal stiffening rod found in all embryoic chordates and in the adults of some of the lower members of this phylum (*Chordata*). It lies between the central nervous system and the alimentary tract (digestive system) and is the axis around which the spinal column develops. As the bony structure forms, the notochord is almost crowded out of existence. In the human body small remnants of it form the nuclei pulposi in the intervertebral disks.

NOVA AND SUPERNOVA. Traditionally, a nova (an early term meaning "new star") has been described as a star that suddenly displays an increasing brilliance and then over a period of time grows fainter. Because of the tremendous distances between the earth and such objects, it is interesting that while some of these events of a cataclysmic nature can be seen by the naked eye from earth as though occurring now, they actually extend far back in time. Most novae that do not attain naked-eye brilliancy are discovered and studied instrumentally.

Supernova 1987A. On the night of 23/24 February 1987, Ian Shelton, working at the University of Toronto Las Campanas Station in northern Chile, discovered the brightest supernova seen since 1604. A separate article on this historic discovery will be found in this encyclopedia: **Supernova 1987A.**

Brightness of Supernova. There is no known procedure by which an estimate of the total number of novae appearing each year can be predicted. It has been estimated that ten or more novae reach a brightness of the ninth stellar magnitude or greater each year, based upon historical records. Between 1900 and 1935, only five novae reached conspicuous brightness. More recent bright novae included V1500 Cyg (Nova 1975 Cyg) and the recurrent nova RS Ophiuci.

In Fig. 1, the light curves of three bright novae of the present century are represented. The ordinate scale of brightness is expressed in stellar magnitude. Since magnitude 6 is the limit of naked-eye visibility, the length of time that each was visible to the naked eye may be determined from the time scale given at the bottom. These curves are characteristic of most novae, with the very rapid rise to maximum and then the relatively slow and irregular decline. Examination of photographic records indicates that novae are not actually new stars at all, but rather are faint stars that suddenly increase in intensity. An increase of ten magnitudes is by no means uncommon, and this represents an increase of light intensity amounting to 10,000-fold. The total emitted energy in one outburst is about 10^{45} ergs.

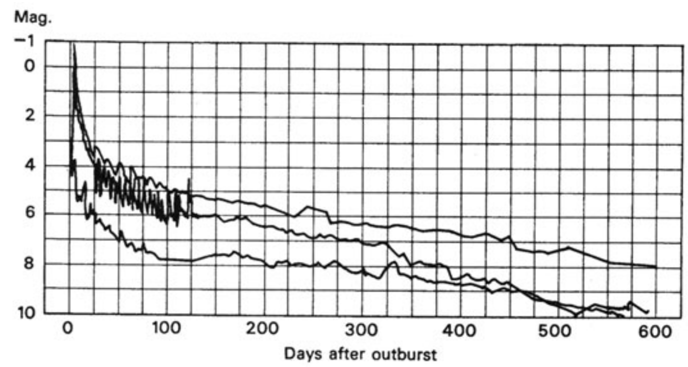


Fig. 1. Light curves of Nova Aquilae, 1918; Nova Persei, 1901; and Nova Geminorum, 1912. They are designated in order of decreasing height. (*Harvard College Observatory Annals.*)

Coupled with the increase in light intensity of a nova is a correspondingly remarkable change in spectral characteristics. Although the spectral changes in different novae vary to a considerable extent, there are certain stages of development that are more or less characteristic of them all. During the period of rise, in the few cases where increasing novae have been detected in time for observation, the stars is of the hot, blue, A-type, with the absorption lines displaced very strongly to the violet. As the star starts to decline, the color changes from white to yellow, and bright lines, particularly of hydrogen and ionized iron, appear. The bright lines then broaden out to bands of irregular structure, which soon completely mask the continuous spectrum of the star. A few days later, dark lines again make their appearance, and these are displaced far to the violet. Soon, bright lines again appear, frequently of the type characteristic of the gaseous nebulae, except that they are broad. Often multiple components, corresponding to separate shells or blobs, are observed. As the brightness of the star further decreases, it eventually settles down to a peculiar O-type spectrum, with bright lines superimposed on a continuous and dark-line absorption spectrum.

Recent infrared observations have revealed that considerable amounts of dust can also be present in the nova ejecta. Radio observations argue for the presence of extensive circumstellar shells pre-existing in the nova environment. Ultraviolet observations show that the period immediately following the nova ejection event is dominated by a strong stellar wind eminating from the still-hot compact star.

Novae have been observed telescopically in some of the exterior galaxies such as the great spiral in Andromeda. Since the distances of some of these objects are at least very approximately known, it is possible to get an approximation to the absolute magnitudes at maxima of the novae observed in them. For these extragalactic novae, we find absolute magnitudes of the order of −4, a value that compares favorably with those determined for the few cases where the distance of a galactic nova is known.

Supernovae fall into two broad categories, mainly on the basis of their light curves. These are:

Type I. These supernovae are around absolute photographic magnitude −18.6 maximum, or more than 200 million times as luminous as the sun. If one of them were placed at the distance of 10 parsecs from Earth, at which distance the sun would be barely visible, the nova would appear 14 times as bright as the full moon does to Earth. The spectra show extremely broad emission bands. The light variations show a rapid rise to maximum, followed at first by a rapid, and later by a slower, decline, usually with a characteristic time scale of 50 to 70 days. See Figs. 2 and 3.

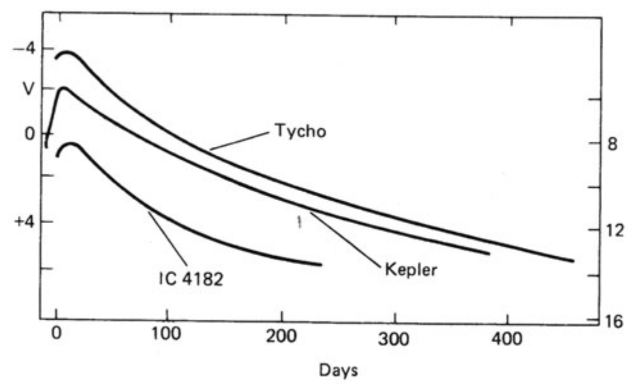


Fig. 2. Light curves of Kepler's and of Tycho's supernovae, as reconstructed by Baade (1945). Scale on right is for the supernova IC 4182. The light curves show that the supernovae of 1572 and 1604 were of Type I. (*After van den Bergh.*)

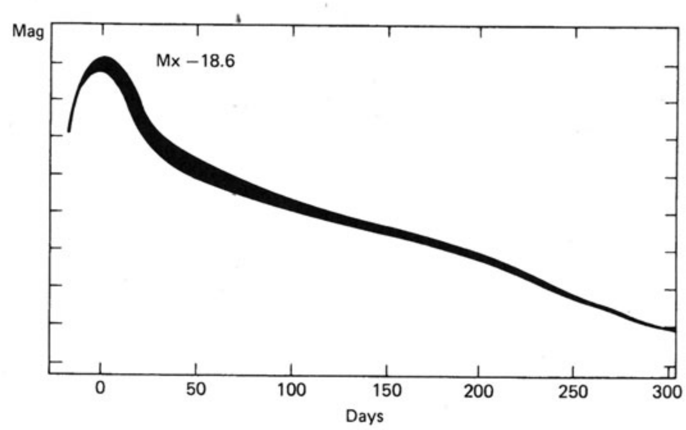


Fig. 3. Composite blue light curve obtained by fitting observations of 38 Type I supernovae. One-magnitude intervals are marked on the ordinates. (*After van den Bergh.*)

Type II. The members of this group reach a maximum luminosity equal to about 20 million suns. After maximum, they fade more slowly at first than do the members of the other group, followed by a more rapid decline after about 100 days. These supernovae (SN) show greater diversity than Type I. See Fig. 4.

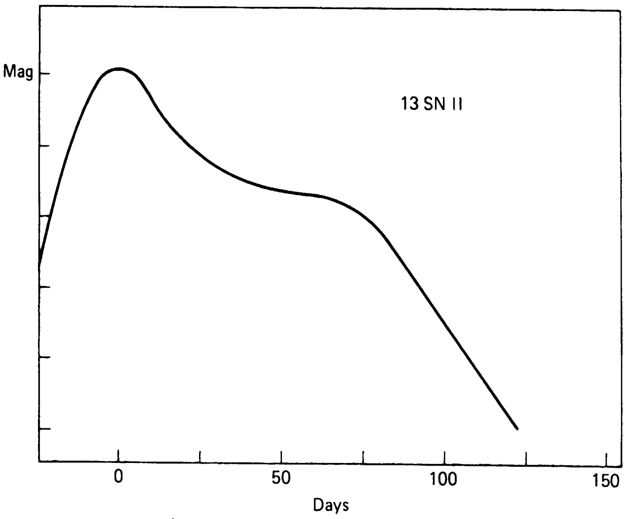


Fig. 4. Composite light curve obtained from 13 Type II supernovae. (*After van den Bergh.*)

Type V. Two extragalactic examples of this class are known. They are more massive than Type II and appear to be associated with very massive (of order 100 $M_\odot$) progenitors.

Historical records indicate that supernovae were observed in AD 185, 1006, 1181, 1572, 1604, and possibly in AD 396. The radio and x-ray source, Cas A, is the remnant of a supernova which may have occurred in the 1670s, although the exact date is not certain. The first extragalactic supernova observed was S And, in M31, which occurred in 1885. Since that time, due to the efforts of groups of many observatories, several hundred extragalactic supernovae have been recorded in a wide class of galaxies. Most galactic supernovae are identified on the basis of their radio and optical remnants, which are caused by the rapid expansion of the ejected matter of the exploded star into the interstellar medium. This shell compresses the magnetic field and generates high-energy electrons which, due to trapping, are swept up with the expanding blast wave. These electrons radiate at radio frequencies, while the shock heats the interstellar gas to temperatures often in excess of a million degrees K, causing the gas to radiate x-rays.

The Crab and Vela supernova remnants have pulsars associated with them, while W 50, a weak optical nebula, has been associated with the peculiar x-ray and optical source SS 433. Few other supernova remnant can, however, be unambiguously linked to pulsars at this writing.

Observations indicate that all novae are members of short-period, low-mass binary systems, in which one of the stars in a white dwarf on which mass lost by the companion (usually a red dwarf or subgiant) is accreting. The slow accumulation of mass eventually causes fusion to begin on the surface of the white dwarf, resulting in the violent ejection of the outer envelope of this hydrogen-rich material. Depending upon the mass of the stars in question, their orbital period and the rate of mass transfer between them, a wide spectrum of explosive behavior can be derived, which appears to cover most of the novae observed thus far. The two types of SN appear to arise from different mechanisms. SNI may be due to the collapse of a white dwarf star, induced by the accretion of matter from a low mass companion in a binary system by a white dwarf near its maximum stable mass. SNII are likely due to the explosion of unstable, massive, recently formed red supergiants. It is suggested that some x-ray pulsars (Cen X-3, HZ Her, for example) represent supernova explosions by members of a binary system which did not disrupt the binary. These are, however, rare exceptions compared with the observed number of supernova remnants.

Steven N. Shore, former Director, Astrophysics Research Center and Associate Professor of Physics,
New Mexico Institute of Mining and Technology,
Socorro, New Mexico, 1989.

Additional Reading

Bethe, H. A., and G. Brown: "How a Supernova Explodes," *Sci. Amer.* 60–68 (May 1985).
Boss, A. P.: "Collapse and Formation of Stars," *Sci. Amer.*, 40–45 (January 1985).
Cassinelli, J. P., et al.: "Central Object of the 30 Doradus Nebua, a Supermassive Star," *Science*, **212**, 1497–1500 (1981).
Chandrasekhar, S.: "On Stars, Their Evolution and Their Stability," *Science*, **226**, 497–505 (1984).
Harris, M. J.: "Short-Lived *s*-Process Gamma-Ray Lines in Type II Supernovae," *Science*, 60 (April 1, 1988).
Hodge, P. W.: "The Universe of Galaxies," W. H. Freeman, New York, 1984.
Iben, I., Jr., and A. Renzini, Editors: "Physical Processes in Red Giants," Reidel, Boston, Massachusetts, 1981.
Kafatos, M. C., and R. B. C. Henry, Editors: "The Crab Nebula and Related Supernova Remnants," Cambridge University Press, New York, 1985.
Kirshner, R. F.: "Supernova—Death of a Star," *National Geographic*, 618 (May 1988).
LoSecco, J. M.: "Detectability of Supernova Neutrinos with an Existing Proton Delay Detector," *Science*, **224**, 56–58 (1983).
News: "Automated Spotting of Supernovae," *Sci. Amer.*, 65 (August 1986).
Peterson, I.: "A Supernova Story in Clay," *Science News*, 397 (June 23, 1990).
Rees, M. J., and R. J. Stoneham, Editors: "Supernovae," Reidel, Boston, Massachusetts, 1982.
Seward, F. D., et al.: "Young Supernova Remnants," *Sci. Amer.*, 88–96 (August 1985).
Spergel, D. N., et al.: "A Simple Model for Neutrino Cooling of the Large Magellanic Cloud Supernova," *Science*, **237**, 1471–1473 (1987).
Staff: "Supernova Yields Cosmic Yardstick," *Science News*, 59 (January 26, 1991).
Waldrop, M. M.: "Supernova 1987A: A Mysterious Stranger," *Science*, **237**, 25–26 (1987).
Waldrop, M. M.: "Sighting of a Supernova," *Science*, **235**, 1143 (1987).
Waldrop, M. M.: "The Supernova 1987A," *Science*, **235**, 1322–1323 (1987).
Waldrop, M. M.: "Feeding the Monster in the Middle," *Science*, 478 (January 27, 1989).
Waldrop, M. M.: "And Now for a Real Crab Nebula . . ." *Science*, 1140 (March 3, 1989).
Weiss, P. L.: "Seeing Supernovas in Galactic *Chimneys*," *Science News*, 133 (September 1, 1990).
Wheeler, J. G., and R. P. Harkness: "Helium-rich Supernovas," *Sci. Amer.*, 50–58 (November 1987).
Williams, R. E.: "The Shells of Novas," *Sci. Amer.*, 120–131 (April 1981).
Wilson, O. C., et al.: "The Activity Cycle of Stars," *Sci. Amer.*, 104–121 (February 1981).

NUBIAN WILD ASS. See **Horses, Asses, and Zebras.**

NUCLEAR FISSION. A type of nuclear reaction in which the compound nucleus splits into two nearly equal parts, rather than ejecting one or a few small nuclear particles, as in most nuclear reactions. Our knowledge of nuclear fission dates back to the mid-1930s when Fermi and his coworkers showed that the number of distinctly different radioactive nuclides that could be induced by neutron bombardment of uranium far exceeded the number expected, unless some previously unknown pattern of isomerism could be found. Furthermore, the radiochemical properties of many of these radio-elements different quite markedly from expectations. For example, both Hahn and Strassman in Germany and Curie and Savitch in France found that certain unknown activities, thought to be radioactive radium, always followed the chemically separated barium fraction rather than the radium fraction. Hahn and Strassman found several other similar examples and were able to show that uranium, when bombarded by neutrons, undergoes what then appeared to be a very unusual nuclear reaction in that the products are radio-elements with about half the atomic number of uranium. These findings were interpreted by Meitner and Frisch as the division of an excited nucleus into nuclei of medium mass, a process that was given the name *nuclear fission*.

The first such process to be extensively studied was fission induced in ^{235}U by thermal neutrons (neutrons with energies of about 0.03 eV). This reaction, symbolically represented by the equation

$$^{235}\mathrm{U} + n \rightarrow {}^{236}\mathrm{U} \rightarrow \text{fission},$$

produces an unstable system which achieves stability by splitting into two large fragments, not by ejecting one or a few small particles.

An individual fission does not produce a unique pair of fragments, but in a large number of such processes, the mass distribution of the fragments can be predicted with reasonable certainty, leading to precitable fission yields. A fission yield, usually expressed as a percentage, describes that fraction of nuclear fission processes that give rise to a specified nuclide or group of isobars. The yields of single nuclides are known as independent yields and those of a set of isobars as mass yields or chain yields. Since two fragments are produced by each fission, the total of all fission yields for a given fission process is 200%. The fission yield curve is different for each mode of induced fission, the most commonly known one being that for thermal neutron induced fission of ^{235}U, shown in Fig. 1. The chemical characteristics of the two fragments vary within limits, so that many elements are formed. Analysis of the fission products shows that most of them are in two mass groups, a "light" group consisting of elements having mass numbers between 85 and 104, and a "heavy" group consisting of elements having mass numbers between 130 and 149. Fragment mass numbers that have been detected range from around A = 70 to around A = 160. The determination of independent yields is made more difficult by the fact that many of the products are highly radioactive and undergo extensive secondary changes, sometimes in extremely short times, a very small fraction of a second.

available neutrons would multiply by factors of two for every generation of fission processes, a very rapid increase. However, some neutrons escape from the region containing the ^{235}U and others are absorbed in nonfission capture processes. The minimum conditions for a self-sustaining chain reaction is that at least one neutron from each nucleus undergoing fission must cause fission of another nucleus, a multiplication factor of one or greater. Maintenance of a chain reaction is essential to the proper functioning of both nuclear weapons and nuclear reactors.

The probability that fission can occur (generally called the cross section for fission) varies widely among different nuclides. Only a few nuclides, such as ^{235}U, have a high probability of undergoing fission when they capture a neutron. In other nuclides, the probability of fission is generally much smaller. As an example, the cross section as a function of incident neutron energy is shown in Fig. 2 for fission of ^{235}U and of ^{238}U. Although fission can be induced in ^{238}U, such a process is possible only if the incident neutron has an energy greater than 1 MeV, whereas neutrons of any energy can induce fission in ^{235}U. The characteristic double hump yield curve of Fig. 1 (asymmetric fission) is common only for low neutron excitation energy and targets consisting of highly fissile elements. For either higher excitation energies or less fissile elements, such as actinium or radium, symmetric fission becomes much more important, creating a triple humped fission-yield curve, shown in Fig. 3. Slightly fissile elements, such as lead and bismuth, or very high excitation energies further emphasize the symmetric mode of fission, also illustrated in Fig. 3. Nuclear fission may be induced by particles other than neutrons, such as alpha particles and photons. In some nuclides, it also occurs spontaneously, although the probability of such occurrence is so low that it has almost no effect on the radioactive decay characteristic of the nuclide.

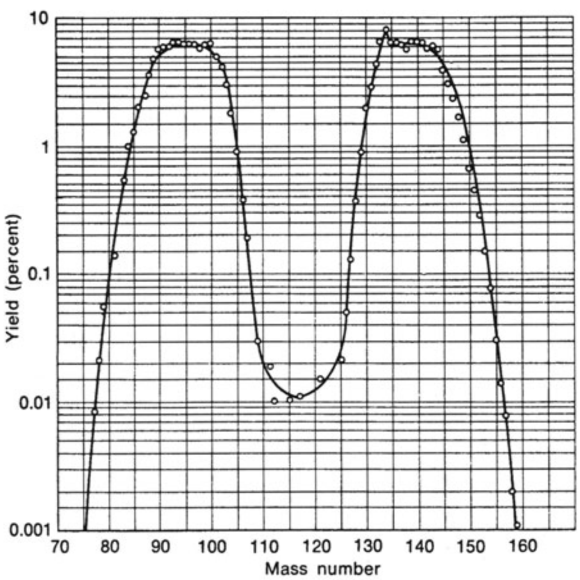


Fig. 1. Mass fission yield curve for $^{235}U + n$ (thermal).

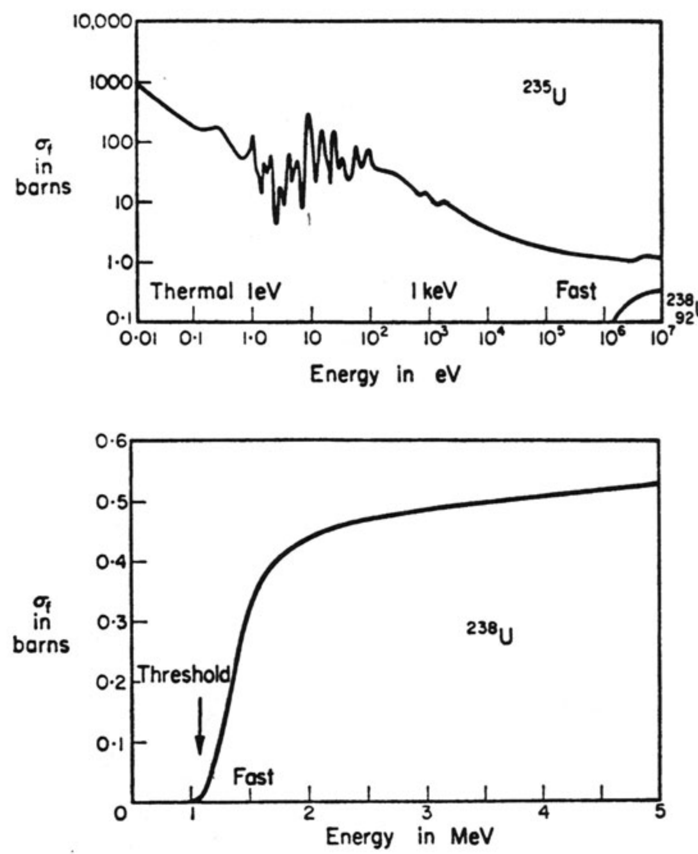


Fig. 2. Fission cross section as function of energy for ^{235}U and ^{238}U.

A most significant aspect of nuclear fission is its great release of energy. The source of this energy is the loss of mass between the initial and final products of the reaction. The total mass of all atoms and nuclear particles produced in a single fission process in less than the original mass of the ^{235}U atom and the neutron that combined with the ^{235}U to induce fission. During fission of ^{235}U, the total energy released because of loss of mass is about 200 MeV. In practical units, the fissioning of 1 gram of ^{235}U yields 24,000 kilowatt-hours of energy.

Another important feature of fission is the presence of neutrons among the reaction products, slightly more than two for each fission of a ^{235}U atom. These neutrons are not an immediate consequence of fission, but are boiled off the original fission products, their release being possible because of the very large amount of available energy. If all these neutrons were captured by other ^{235}U nuclides, the number of

Nuclear fission has generally been explained theoretically in terms of the liquid-drop model of the nucleus. In this model, the incident neutron combines with the target nucleus to form a compound nucleus at a high excitation energy. A small part of this excitation energy can be attributed to the kinetic energy of the incident neutron, but most of it

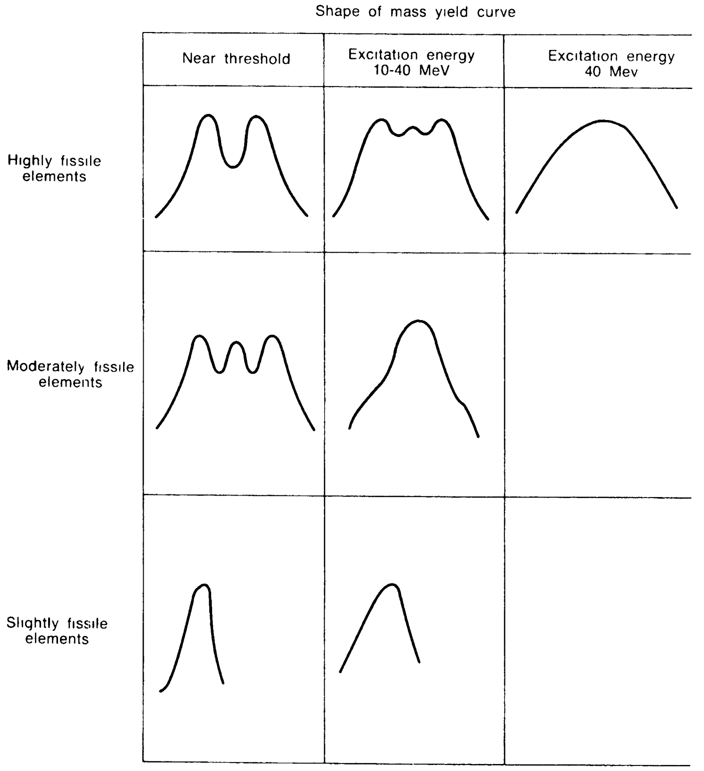


Fig. 3. Mass fission yield curves as function of excitation energy and degree of fission probability.

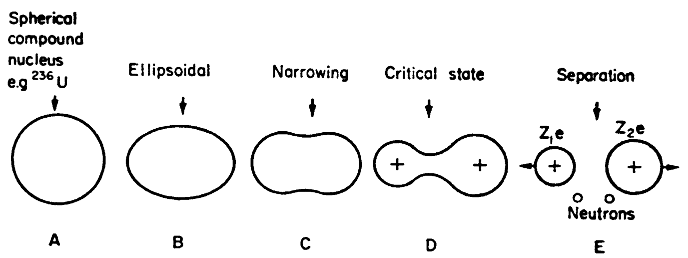


Fig. 4. Fission mechanism according to liquid-drop model of the nucleus.

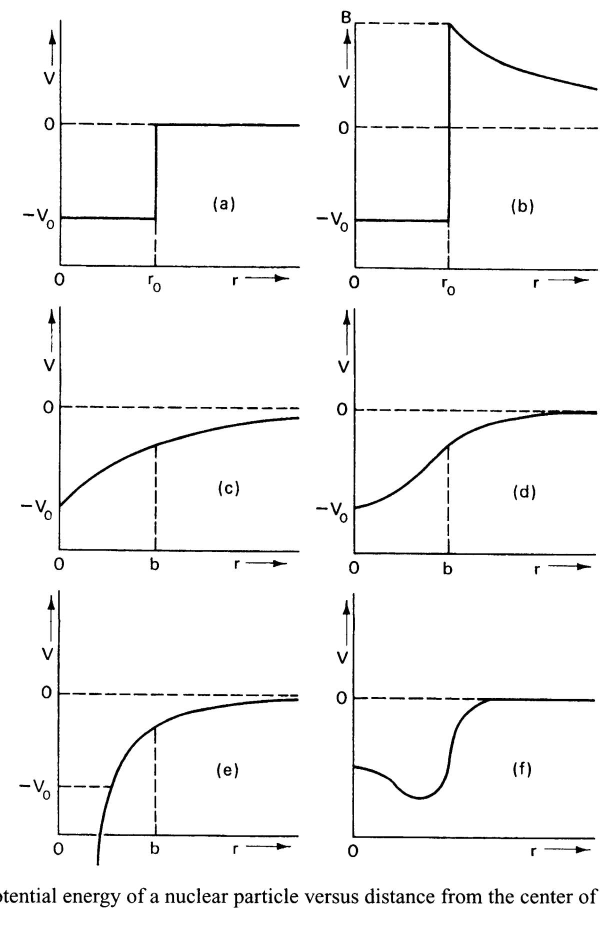
usually comes from the binding energy of the incident neutron. This added energy initiates oscillations in the drop, which then sometimes assumes an elongated shape, similar to B in Fig. 4. If oscillations become sufficiently violent that a form similar to D is reached, fissioning (form E) becomes inevitable, since the positive charge at the two ends of the dumbbell-shaped nucleus then produces an electrostatic repulsive force greater than the attractive nuclear force holding the neck of the dumbbell together. The reason for asymmetric fission is not clearly understood. The liquid drop model predicts symmetric fission. Most people believe that asymmetric fission results because of the effects of the closed shells of the nucleus. See also **Nuclear Power; Nuclear Structure.**

C. Sharp Cook, Professor of Physics,
The University of Texas at El Paso, El Paso, Texas.

NUCLEAR FORCES. Strong, short-range, attractive forces that interact between the individual nucleons of an atomic nucleus. Unfortunately, despite several decades of research, a clear and unambiguous description cannot be given for the forces that hold individual protons and neutrons together in an atomic nucleus. Unlike the electrostatic force that holds electrons in an atom, no equation can be written that completely describes the nature of the force that holds an atomic nucleus together, or the nature of its associated potential energy. A description of the detailed structure of a nucleus cannot, therefore, be derived directly from calculations based on knowledge of nuclear forces. Instead, detailed knowledge of the structure of atomic nuclei has been derived from nuclear models. These models have been constructed by using results from other fields of physical science which display the same or similar characteristics as those observed in nuclear reactions and in radioactive decay. From such analogies, construction of a partial description of nuclear structure and of the nature of nuclear forces has been possible.

Because of the unknown characteristic of nuclear forces, many different suppositions have been made, using available experimental evidence, regarding the nature of the potential energy V of a nuclear particle as a function of its position in the field of a nucleus, or of another nuclear particle. To a first approximation, the nuclear potential is assumed to be spherically symmetric, such as V is a function only of the distance r from the center of the field, thus being the same in all directions, and is representable by a curve as in accompanying curves (a) to (f).

A *potential well* is the name given to a region in which a minimum in the potential is formed; it results from attractive forces. A *potential barrier* is the name given to a region in which there is a maximum in the potential; it results from repulsive forces, either alone or in combination with attractive forces. Some central potentials commonly used as approximations to nuclear potentials are illustrated in the curves. Curve (a) shows a square well potential, which has a constant negative value $-V_0$ for $r \leq r_0$ and zero value for $r \geq r_0$. When this curve represents the potential between two nucleons, r_0 is called the *range of nuclear forces*; when it represents the potential of a nucleus, as this nucleus interacts with an individual nucleon, r_0 is called the nuclear radius. Curve (b) shows a square well potential for $r \leq r_0$ with a Coulomb potential resulting from repulsive electrostatic forces, for $r > r_0$. The resulting barrier is called a *Coulomb barrier*, and the maximum energy b is called the *barrier height*. Such a potential approximates that of a positively charged particle in the field of a nucleus, and is often used in the theory of alpha particle distintegration and nuclear reactions. Curve (c) shows an exponential well, $V = V_0e^{-r/b}$; curve (d) shows a Gaussian well, $V = V_0e^{-r/b2}$. Curve (e) shows a Yukawa poten-



Potential energy of a nuclear particle versus distance from the center of the field.

tial, $V = -(V_0/r)e^{-r/b}$ used in the meson theory of nuclear forces for the interaction between two nucleons; and (f) shows a wine-bottle potential, characterized by a low central elevation. If a high central elevation is present, the resulting barrier is called a *central barrier*, or *repulsive core*.

Although the predominant part of the nuclear potential is the part described above that is derived from the central force produced by the average effects of all other nucleons in the system on the individual nucleon under observation, evidence exists that nonsymmetric tensor and spin-orbit coupling terms must be included in the description of the nuclear potential. These are derived from a tensor force resulting from a coupling between individual pairs of nucleons and from the coupling between spin and orbital angular moments of the individual nucleus, as described by the shell model of the nucleus.

A considerable amount of evidence indicates that nuclear forces are charge-independent, i.e., the neutron-neutron, neutron-proton, and proton-proton forces are identical. The meson theory of nuclear forces, originated by Yukawa, postulates the atomic nucleus being held together by an exchange force in which particles, now called mesons, are exchanged between individual nucleons within the nucleus.

C. Sharp Cook, Professor of Physics,
The University of Texas at El Paso, El Paso, Texas.

NUCLEAR MAGNETIC MOMENT. An electrically charged particle of finite size that possesses angular momentum, acts like a small magnet and thus possesses a magnetic moment. For atomic nuclei, the magnitude of this moment is within a range of values between zero and a few nuclear magnetons.

NUCLEAR MAGNETIC RESONANCE (NMR) AND MAGNETIC RESONANCE IMAGING (MRI). Since the discovery of electron spin resonance (ESR) by Zavoiski in 1945 and the codiscovery of nuclear magnetic resonance (NMR) in 1946 by Purcell, Pound, and Torrey (Harvard University) and Bloch, Hansen, and Packard (Stanford University), the field of magnetic resonance has reached a high degree of sophistication and versatility.

Soon after its discovery, the principle was applied by physicists to a variety of research programs. Chemists also became interested in NMR because they recognized its potential for elucidating structure and dynamics in pure and applied chemistry. NMR has gained exceptional acceptance over the intervening years as a tool for yielding details of molecular structure, such interests coinciding with other developments in molecular biology. The use of NMR in medical research dates back about 40 years, commencing with proton NMR measurements on cells and organs of both humans and laboratory animals. During the past 20 years, the applications of NMR to clinical medicine have expanded rapidly. NMR permits the analysis of protein structures, for which the former X-ray methods were inadequate. By the late 1980s, the structures of some 50 important proteins had been determined. NMR allows the investigation of proteins and peptides and other macromolecules in an aqueous medium. Consequently, the effects of pH (hydrogen ion concentration) can be observed and the binding of water at the interior and surface of proteins can be determined.

The molecular weights of proteins studied thus far range from about 5000 to 15,000. The use of three-dimensional Fourier transform NMR may ultimately permit the study of macromolecules up to a molecular weight of 40,000.

In analytical chemistry as of the early 1990s, NMR is routinely used to study: (1) polymers, polymer networks, and copolymers, and (2) asphalt, bitumens, tars, and pitches, among numerous examples that could be cited. Refer to the Murray reference listed.

Commencing with clinical medicine (circa 1981), which led to the development of NMR and magnetic resonance imaging (MRI) for medical research and diagnostics, the medical applications have captured the "headlines," particularly since the early 1990s. It is interesting to note that in 1981 there were only two MRI devices in the United States. By 1987, the installations had reached about 600 in the United States and about 100 MRI facilities in Europe. As of 1994, these figures have expanded into several hundred installations, representing a tremendous investment by the health care field.

Fundamentals of the Technology

Discovery of the principles of ESR and NMR led to the development of several subtechnologies.

Electron Spin Resonance (ESR). Fundamentally, ESR spectrometers consist of three main parts: (1) a magnet with corresponding power supply to provide a steady dc magnetic field at about 3300 G; (2) a microwave bridge capable of producing an oscillating electromagnetic field at a frequency of about 0.5 GHz (*X* band), which is coupled via waveguide to a high-*Q* microwave reflection cavity; and (3) the associated signal detection including dc field modulation, amplification, and display systems. See Fig. 1.

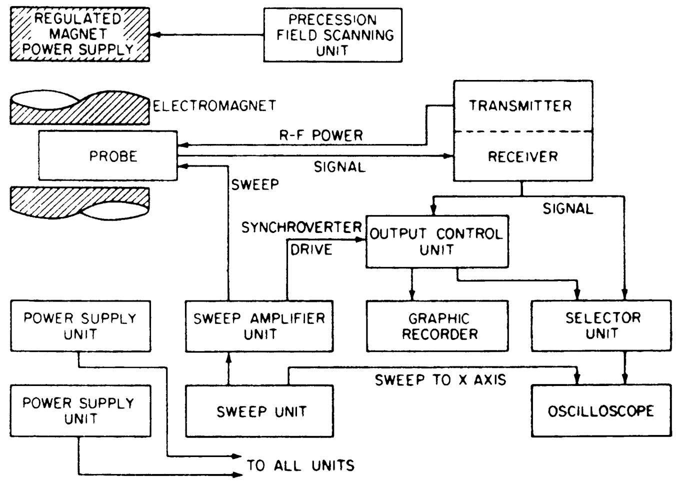


Fig. 1. Schematic of ESR spectrometer.

Observation of ESR from a particular sample is contingent upon the presence of a macroscopic spin magnetic moment $\bar{\mu}$; i.e., the sample under investigation must contain some minimum number of unpaired electron spins. Upon insertion into the cavity, the sample is subjected to the dc magnetic field H_0, and the unpaired electrons align themselves both parallel (high energy) and antiparallel (low energy) to H_0. The ratio of the spin populations of the high to low energy states is given by the Botzman distribution, $e^{-h\nu/kT}$, where $h\nu$ is the energy difference of the two states. Because of the torque on $\bar{\mu}$ produced by H_0, the spin magnetic moment will precess about H_0, at the Larmor frequency $\omega = \gamma_e H_0$, where γ_e is the gyromagnetic ratio of the electron.

Analysis is accomplished by sweeping the magnetic field until the Larmor frequency of the spin system is identical to the fixed frequency of the oscillating microwave field emanating from the bridge. When the two frequencies are coincident, a net microwave energy will be absorbed by the spin system from the oscillating field because of the excess spin population in the lower energy state. If energy continues to be absorbed, the spin populations will equalize and saturation will occur—no *net* microwave energy will be absorbed and the signal will disappear. To avoid this situation the spin system interacts with its surroundings (lattice) and transfers the absorbed microwave energy to the lattice in some interval called the spin-lattice relaxation time.

This phenomenon, known as relaxation, acts to restore the spin system to its original Botzman distribution of populations. At equilibrium, microwave energy is being absorbed by the spin system and then transferred to the lattice. This process is monitored via microwave rectification of the reflected cavity signal by a diode detector, preamplification, and lock-in phase detection at the dc field modulation frequency. The signal which appears on the recorder may be a single line from an unpaired electron or a group of lines. The latter are caused by neighboring nuclei with nonzero nuclear spin (hyperfine structure) and by surrounding electric field gradients (quadrupole interactions).

In this manner the unpaired electron spins may be used as a probe for analysis of their immediate microscopic surroundings. Typical instrument sensitivity is such that approximately 10^{-11} mole of a paramagnetic species can be detected, but generally concentrations of 10^{-4} mole give optimum ESR spectra.

Nuclear Magnetic Resonance (NMR). This technique is essentially based on the same principle as ESR, but NMR is capable of detecting nuclei (MHz) instead of electrons (GHz). (Lack of a standardized nomenclature has resulted in numerous modifiers in connection with magnetic resonance instrumentation—electron, proton, nuclear, etc., plus application-related terms, such as silicon-29, oxygen-17, ^{13}C, ^{31}P NMR, etc.)

In nuclear magnetic resonance spectroscopy, a nucleus possessing a magnetic moment when placed in a homogeneous magnetic field will precess about the field axis at a rate which is dependent upon the strength of the field. See Fig. 2. If these nuclei are then brought into contact with an oscillating electromagnetic field of the same frequency as the Larmor precessional frequency of the nuclei, energy will be absorbed by the nuclei spin system from the oscillating ratio frequency field. As stated above, the net energy absorption is proportional to the Boltzman distribution of spin populations and the effective nuclear relaxation times.

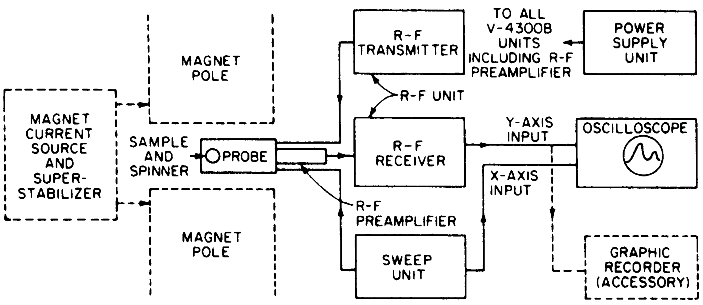


Fig. 2. Schematic of NMR spectrometer.

Different nuclei will have different precessional frequencies and therefore at a particular field will absorb energy at certain characteristic radio frequencies. Also, nuclei of the same nuclear species (such as hydrogen) will absorb energy at slightly different frequencies, dependent upon their molecular environment. The latter observation makes possible an entire subfield of NMR spectroscopy, termed high-resolution NMR, which is an appropriate method for determining chemical structure and identifying and measuring similar nuclei in two or more different compounds (mixtures).

The major components of a magnetic resonance spectrometer are: (1) a magnet capable of producing a very strong homogeneous field which may be continuously varied over a very small range; (2) a low-power radio frequency oscillator which supplies rf power to a small transmitter coil surrounding the sample; (3) a small receiver coil which also surrounds the sample (but is orthogonal with respect to the transmitter), and feels (4) a sensitive radio receiver (tuned to the same frequency as the transmitter) capable of amplifying any signal which might be induced in the receiver; and (5) a recorder of oscilloscope which can display the resulting spectra.

Various decoupling techniques are used to simplify complex NMR spectra. For example, in a simple two-spin system (homonuclear) giving rise to four NMR lines it is possible to saturate a nucleus at a particular rf frequency and collapse the remaining doublet to a single NMR line. This occurs because the second nucleus sees an averaged interaction from the first rather than two distinct interactions from spins in the high and low energy states. Noise decoupling is predominantly used to decouple different nuclei—for example, hydrogen nuclei C^{13} spectra (heternuclear). Here the entire hydrogen spectrum is saturated over a range of frequencies (noise) leaving behind the much simplified C^{13} spectrum.

In internuclear double resonance one nuclear line at a given frequency is observed while a second frequency is swept through the remainder of the NMR spectrum. All nuclei which are in some way coupled to the line being observed will enhance or de-enhance the latter as the second frequency passes through the resonance condition. Also, the Fourier transform technique has nicely complimented NMR. Any complex waveform can be converted to a spectrum of frequencies by Fourier transformation.

In NMR, the waveform is a superposition of a set of nuclear precession frequencies with amplitudes decaying due to relaxation and field inhomogeneity. The transformation may be carried out by analog means (spectrum analyzer) or on a small dedicated laboratory computer. The latter appears to be the most convenient solution. Since the free induction decay signal decays with time, whereas the instrumental noise remains constant, the noise content is higher in the tail of the transient signal, and it is possible to improve the overall signal-to-noise ratio by weighting the transient signal with an exponentially decaying function of time. The shorter the time constant of this exponential, the greater the improvement in sensitivity, but this increases the linewidths of the transformed spectrum. The reversal of this procedure can enhance resolution at the cost of sensitivity.

High-Resolution Nuclear Magnetic Resonance of Solids. Important developments in solid-sample NMR techniques of the early 1980s have made NMR of significant interest as a tool for characterizing solid samples—as it has been in the past for the study of liquids. As observed by Maciel, the development of line-narrowing techniques, such as magic-angle spinning (MAS) and high-power decoupling, has led to powerful high-resolution NMR for studying solids. In favorable cases (for example, where high abundances of protons are present) cross polarization (CP) provides a means of circumventing the time hurdle caused by inefficient spin-lattice relaxation in many solids. Combining the CP and MAS approaches for carbon-13 with proton decoupling has become a popular and routine experiment for organic solids. For many nuclides with spin quantum number $1 > \frac{1}{2}$, the central nuclear magnetic resonance transition can be used in high-resolution experiments that involve rapid sample spinning. A continuing stream of other advances in NMR technology bodes well for the characterization of solids by a wide range of nuclides. The complexities of this topic unfortunately are beyond the scope of this encyclopedia.

It is interesting to note that several of the concepts for improving NMR technology, as listed by Levy and Craik, in 1988, already have been partially or fully achieved: (1) two-dimensional Fourier transform (FT NMR); (2) high-resolution NMR in solids; (3) new types of pulse sequences; (4) chemically induced dynamic nuclear polarization; (5) multiple quantum NMR; and (6) NMR imaging (MRI).

Two-Dimensional NMR. Bax and Lerner report on how two-dimensional Fourier transform pulse NMR (2-D FT NMR) has extended the range of applications of NMR spectroscopy into the area of large, complex molecules, such as DNA and proteins. Great spectral simplification can be obtained by spreading the conventional one-dimensional NMR spectrum in two independent frequency dimensions, thus removing spectral overlap, facilitating spectral assignment, and providing additional information. Conformational information related to interproton distances is available from resonance intensities in certain types of two-dimensional experiments. Two-dimensional NMR spectroscopy also has been applied to the study of ^{13}C and ^{15}N to provide connectivity information and greatly improving the sensitivity of these determinations. A traditional NMR spectrum of a sugar and a 2D spectrum are contrasted in Fig. 3.

Correlative spectroscopy (COSY) is an approach that involves correlating groups believed to be coupled to each other and to prove that this coupling does exist. *Spin-echo correlation spectroscopy* (SECSY) is a variation of the COSY technique. Sometimes called *J-resolved spectroscopy*, it is a technique that allows a separation of the chemical shift of a nucleus from the coupling to other nuclei. This simplifies the spectrum and permits one to assign each resonance to a specific nucleus. Sometimes referred to as *nuclear Overhauser effect* (NOESY), this is a technique that makes it possible to determine distances between nonadjacent residues in a peptide chain. Another application, *multiple quantum transitions*, takes advantage of the fact that molecules in a sample are forced to absorb or emit several quanta of energy at one time. For example, the technique can be used to determine which carbon atoms in a molecule are connected to other specific atoms in the molecule.

Magnetic Resonance Imaging (MRI). In 1973, Lauterbur added a new aspect to NMR basics, namely that of image formation based upon NMR principles. This led to NMR imaging, now commonly referred to as magnetic resonance imaging (MRI).

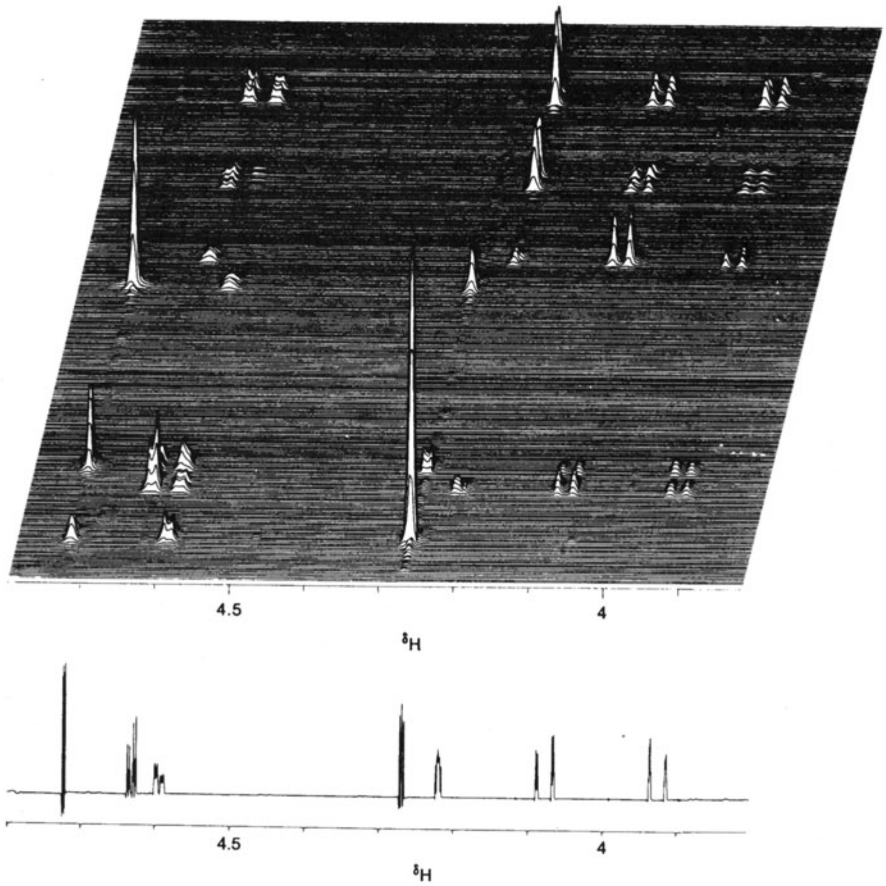


Fig. 3. (*Top*) A two-dimensional nuclear magnetic resonance spectrum of a sugar; (*Bottom*) conventional NMR spectrum of same sugar. The two-dimensional spectrum also can be plotted as a contour map with intensities denoted by color. (*JEOL Inc.*)

The source of the imaging photons is not a Roentgen tube, nor an unstable isotope, nor an electron storage ring, but rather it is a radio transmitter. The commonly referred *radio photon* radiation allows photons to pass into the tissues under study, where they are absorbed by nuclear particles rather than by electrons. As pointed out by Rubenstein, the nucleons are momentarily excited by the process and then return to their resting state by emitting photons of the same or nearly the same energy that they had absorbed. Because the nuclei of the different chemical elements absorb radio photons of different frequencies, it is possible to use the method to detect the presence of a single element in a sample. When protons are irradiated by *radio* photons of a frequency that precisely matches their own precessional frequency, resonance occurs. The resonant frequency (Larmor frequency) is determined by the natural rotational velocity (gyromagnetic ratio) characteristic of each species of atomic nucleus and by the strength of the applied static magnetic field, as expressed by: Larmor frequency = Magnetic field × Gyromagnetic ratio. See Fig. 4.

Magnetic resonance imaging is well suited for the imaging of soft tissues. It has been found particularly effective in (1) studying skeletal musculature, notably in the male and female pelvic regions, (2) delineating tumor development and assisting in planning surgery or therapy, and (3) precisely locating tumors, particularly in the brain where, with time, the evolution of hematomas can be studied. For clinical work, MRI installations may be used in two staff shifts per day, holidays, and weekends.

The magnetic fields used are usually 1.5 to 2 teslas (15,000 to 20,000 gauss). Barriers to using higher magnetic fields are magnet cost, additional expense for building the site, and notation by some persons (volunteers) to discomfort when over 4 teslas are used, which also causes the patient to move.

In 1986, Fossel and McDonough (Beth Israel Hospital, Boston) proposed that proton NMR spectroscopy of human plasma possibly could be a "potentially valuable approach to the detection of cancer and the monitoring of therapy." Subsequent investigations through 1990 failed to develop a convincing correlation.

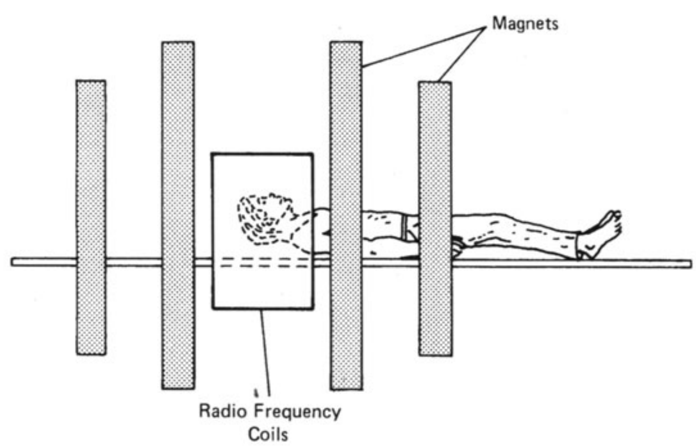


Fig. 4. General arrangement of MRI equipment in a medical setting.

Traditional MRI of the human body relies mainly on the detection of the most abundant type of nuclei, the hydrogens in water, and, to some extent, fat.

As pointed out by C. Moonen (National Institutes of Health), "For discrimination of healthy and diseased tissues, adequate contrast is essential. Such contrast depends not only on differences in water concentration, but also on the NMR relaxations, which, in turn, are related to local mobilities and interactions."

Magnetic resonance imaging, in addition to providing detailed information about the macroscopic structure and anatomy, also permits the noninvasive spatial evaluation of various biophysical and biochemical processes in living systems. These include the motion of water in processes, such as vascular flow, capillary flow, diffusion, and exchange. Further, the concentrations of various metabolites can be determined for the assessment of regional regulation of metabolism. In the scholarly Moonen paper, examples are given of flow imaging, diffusion imaging, and the imaging of tissue perfusion, of exchange, and of metabolites. These aspects of MRI imaging sometimes are referred to as *functional MRI.*

Imaging of the central nervous system by MRI technology is of great interest. Advantages of MRI include its sensitivity to soft tissue, the contrast between gray and white matter, the paucity of signals from the skull, and the availability of coronal, sagittal, and transverse sections. The procedure has been used to detect posterior fossa tumors, such as acoustic neuromas, and pituitary and parasellar tumors, orbital tumors, multiple sclerosis, and a number of other lesions in the craniovertebral junction, and of the cord and spine. MRI technology, according to a number of authorities, has large potential for use in *noninvasive* cardiological studies, particularly of blood flow imaging. Contrast agents are not required because nuclei in rapidly flowing blood move out of the volume of interest during the interval required for application of the rf pulse and of gradient magnetic fields. The parameters that influence MRI signals at various flow rates are being studied.

In addition to hydrogen atoms, MRI can create, for example, ^{31}P images which are excellent indicators of energy metabolism. ^{23}Na images reflect extracellular and intracellular fluid fluxes.

Echo-Planar Imaging. A major problem with MRI in the past has been the long data acquisition times (up to several minutes). Consequently, MR images are subject to so-called *motional arfifacts,* caused by physiological motions (heartbeat, blood flow, bowel peristalsis, breathing) as well as by voluntary movements in severely ill and uncooperative patients, including children. Echo-planar imaging (EPI) permits faster scan times, thus effectively reducing imaging to a fraction of a second as compared with minutes. Although a technical description of how EPI is implemented is too complex for coverage here, echo-planar imaging uses only one nuclear spin excitation per image.

EPI has broadened the use of MRI to include the evaluation of cardiac function in real time, mapping of organ blood pool and perfusion, functional imaging of the central nervous system, depiction of blood and cerebrospinal fluid flow dynamics, and motion picture imaging of the mobile fetus in utero. EPI also has the practical advantages of increasing patient throughput at a lower cost per MRI examination. With these advantages, it is expected that EPI will become an established tool for early diagnosis of some common and potentially treatable diseases, such as ischemic heart disease and stroke.

Comparison with Ultrasonography. The health care community is aware of the high costs of MRI and continues to seek other techniques that may be effective at lower cost. One such comparison was made in connection with prostate cancer. The approach to treatment varies and depends on the extent of cancer at the time of diagnosis.

In a specific comparative study over a period of 15 months, 230 patients were evaluated with identical imaging techniques. It was found that MRI correctly staged 77% of cases of advanced disease and 57% of cases of localized disease. The corresponding figures for ultrasonography were 66% and 46%. MRI identified only 60% of all malignant tumors measuring more than 5 mm on pathological analysis, while ultrasonography identified only 59%. The study concluded, "The MRI and ultrasonography equipment is not highly accurate in staging early prostate cancer, mainly because neither technique has the ability to identify microscopic spread of disease."

Superconducting Quantum Interference Device (SQUID). These devices are sensitive detectors of magnetic fields. Low-temperature superconductors have been used in the past to sense weak magnetic signals from the brain for medical diagnosis. Such devices, however, required cooling with liquid helium and thus have resulted in costly, unwieldy apparatus. Research is now underway toward using higher-temperature superconductors, such as $YBa_2Cu_3O_7$. Although the higher-temperature semiconductor still requires refrigeration, this can be accomplished with liquid nitrogen, thus resulting in a less costly, simpler, and more portable detecting device.

Additional Reading

Abelson, P. H.: "New Horizons in Medicine," *Science*, 1109 (November 25, 1988).

Ancreasen, N. C.: "Brain Imaging: Applications in Psychiatry," *Science*, 1381 (March 18, 1988).

Bain, L.: "MRI—Safety Issues Stimulate Concern," *Science*, 1245 (May 31, 1991).

Bradley, W. G., and G. Bydder: "MRI Atlas of the Brain," Raven Press, New York, 1990.

Edelman, R. R., and J. R. Hesselink, Editors: "MRI: Clinical Magnetic Resonance Imaging," W. B. Saunders, Philadelphia, Pennsylvania, 1990.

Hari, R., and O. V. Lounasmaa: "Recording and Interpreting Cerebral Magnetic Fields," *Science*, 432 (April 28, 1989).

Huk, W. J., Godemann, G., and G. Friedman, Editors: "MRI of Central Nervous System Diseases," Springer-Verlag, New York, 1990.

Kang, H. S., and D. Resnick: "MRI of the Extremities," W. B. Saunders, Philadelphia, 1991.

Kazner, E., et al., Editors: "Computed Tomography and MR Tomography of Intercranial Tumors: A Clinical Perspective," Springer-Verlag, New York, 1989.

Kirkwood, J. R.: "Essentials of Neuroimaging," Churchill-Livingstone, New York, 1990.

Moonen, C. T., et al.: "Functional Magnetic Resonance Imaging in Medicine and Physiology," *Science*, 53 (October 6, 1990).

Pool, R.: "Putting SQUIDs to Work," *Science*, 862 (August 24, 1990).

Randal, J.: "NMR: The Best Thing Since X-Rays," *Technology Review (MIT)*, 59 (January 1988).

Rifkin, M. D., et al.: "Comparison of Magnetic Resonance Imaging and Ultrasonography in Staging Early Prostate Cancer," *New Eng. J. Med.*, 623 (September 6, 1990).

Runge, V. M., Editor: "Clinical Magnetic Resonance Imaging," J. B. Lippincott, Philadelphia, Pennsylvania, 1990.

Shulman, R.: "NMR—Another Cancer-Test Disappointment," *New Eng. J. Med.*, 1002 (April 5, 1990).

Sochurek, H., and P. Miller: "Medicine's New Vision," *National Geographic*, 2 (January 1987).

Stehling, M. K., Turner, R., and P. Mansfield: "Echo-Planar Imaging in a Fraction of a Second," *Science*, 43 (October 4, 1991).

Sutton, D., and J. W. R. Young, Editors: "A Short Textbook of Clinical Imaging," Springer-Verlag, New York, 1990.

NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY. See **Magnetic Resonance Spectroscopy.**

NUCLEAR MAGNETON. A unit of magnetic moment used in atomic and nuclear physics, defined as

$$\mu_N = eh/4\pi M = 5.050 \times 10^{-27} \text{ ampere-meters}^2,$$

in which e is the electronic charge (1.602×10^{-19} c), h is Planck's constant, and M is the rest mass of the proton. This unit is derived for nuclear particles by analogy with the Bohr magneton, which is applicable to the magnetic moment associated with electron orbital angular momentum in an atom. No real nuclear particle, however, has a magnetic moment of exactly one nuclear magneton. The magnetic moment of the proton is +2.793 nuclear magnetons, and of the neutron −1.913 nuclear magnetons.

NUCLEAR POTENTIAL. The potential energy V of a nuclear particle as a function of its position in the field of a nucleus or of another nuclear particle. A central potential is one that is spherically symmetric, so that V is a function only of the distance r of the particle from the center of force. A noncentral potential, on the other hand, is one that is not spherically symmetrical, or that depends upon the relative directions of the angular momenta associated with the particle and the center of force as well as upon the distance r. A negative potential corresponds to an attractive force, while a positive potential corresponds to a repulsive force.

Although the expression can certainly be applied to the problem of nuclear forces, the usual meaning of a nuclear potential refers to the interaction of a nucleon (neutron or proton) with a complex nucleus. Although the potential energy of a single nucleon inside a nucleus is clearly a rapidly varying function of position and time (since it represents the interaction with a large number of closely packed, fast-moving particles), one may nevertheless speak of the average potential energy, and regard this as a smoothly varying function. For a neutron, the nuclear potential is essentially negative inside the nucleus, rising rapidly to zero outside the nuclear radius R. For a proton, the long-range electrostatic repulsion must, of course, be added. Owing to the Pauli exclusion principle, and to the exchange nature of nuclear forces, however, such a potential cannot in general be regarded simply as a function of position, $V = V(r)$; it depends in addition upon the momentum of the particle, which in quantum mechanics does not commute with the position. Hence, the potential must be regarded as a nondiagonal matrix operator $V = \langle r|V|r'\rangle$ in configuration space, or a similar operator in momentum space.

Although the concept of a nuclear potential in this latter sense cannot be defined in a precise way, it has nevertheless been useful, both qualitatively and quantitatively in the investigations of nuclear structure and nuclear reactions. It has been of particular usefulness in the optical model of nuclear reactions.

NUCLEAR POWER TECHNOLOGY. After a terse review of the physics and chemistry of nuclear fission and the nature of these reactions, some of the design features and operating parameters of the four families of reactors currently installed are described. These units, some of which are approaching their retirement within the next decade or so, are, of course, the starting basis for the design of next-generation reactors, which will appear in substantial numbers prior to the end of the 20th century. Design improvements that constitute these forthcoming systems are described in terms of their status as of early 1994. The topic of radioactive waste handling is summarized. Based upon available information, nuclear power technology in the United States as well as in Canada, France, Japan, the U.K., and other leading industrialized nations is covered.

The first nuclear power plant was installed at Shippingport, Pennsylvania in 1957 and, after serving as a test facility for several years, was dismantled because of "old age." The plant had a capacity of 60,000 kilowatts. Indeed, the plant was extremely small by comparison with hundreds of units installed today in the United States and other major countries of the world.

NATURE OF NUCLEAR FISSION REACTIONS

The energy of a nuclear fission reaction can be computed from the change in mass between reactants and products according to Einstein's law:

$$\Delta E = \Delta mc^2$$

where E is the energy in ergs, m is mass in grams, and c is the velocity of light in centimeters per second. For example, the mass difference in this equation is $\Delta m = 0.2058$ amu (atomic mass units). Therefore, $\Delta E = 931$ MeV/amu $\times$ 0.2058 amu = 191.6 MeV. The average amount of energy released in the various fission reactions is about 200 MeV (million electron volts). This energy is distributed in the fission process as:

	MeV
Kinetic energy of fission fragments	165
Radioactive-decay energy	23
Kinetic energy of neutrons	5
Prompt gamma-ray energy	7

The energy of a chemical reaction, approximately 3–4 eV, is dramatically lower than that of a nuclear reaction. Hence, the fission of ^{235}U yields 2.5 million times as much energy as the combustion of the same weight of carbon.

The importance of fission in energy (power) production lies in two facts: (1) an exceedingly large amount of energy is released in the fission reaction; and (2) the production of excess neutrons permits a chain reaction. These two circumstances make it possible to design nuclear reactors in which self-sustaining reactions occur with the continuous release of energy. Although described later, it may be pointed out here that nuclear fission is not the only energy-releasing nuclear reaction. The *fusion* of light nuclides, like hydrogen, into heavier elements is also an energy-producing process.

The heat generated in nuclear power plants is transferred to a working fluid and from this point on the nuclear power plant and the conventional fossil-fueled power plant are essentially similar.

Fission Reaction. In nuclear fission, the nucleus of a heavy atom is split into two or more fragments. The reaction is initiated by the absorption of a neutron. A typical reaction is

$$^{235}_{92}\text{U} + ^{1}_{0}n \rightarrow ^{137}_{56}\text{Ba} + ^{97}_{36}\text{Kr} + 2\,^{1}_{0}n + \Delta E$$

In this reaction, a ^{235}U atom absorbs one neutron, becomes unstable, and subsequently fissions into two fission fragments plus two neutrons. This is just one of the many ways in which ^{235}U might fission. The number of neutrons produced in a fission reaction is usually 2 or 3. The excess neutrons produced by the fission reaction provide the means of self-sustaining the chain reaction. Nuclides including ^{233}U, ^{235}U, and ^{239}Pu, which are fissionable by neutrons of all energies, are termed *fissile* nuclides.

Nuclear Fuels. There are two broad categories of nuclear fuels: (1) the fissile nuclides previously mentioned; and (2) the fissionable nuclides, ^{232}Th (thorium) and ^{238}U. Thermal reactors use fissile nuclides as fuel, while *fast reactors* are designed to burn fissionable materials. In fast reactors, only a small portion of the ^{232}Th and ^{238}U are fissioned directly. A larger portion of these materials is converted into ^{235}U and ^{239}Pu, respectively, through neutron absorption. Thus, this type of reactor not only consumes fuel, but also produces (breeds) new fuel material. Hence, the term *breeder reactor* is used for reactors designed to take advantage of this phenomenon. Breeding is possible in thermal reactors also, but to a lesser extent. The fuel material in a fast reactor must contain a significant amount (about 10%) of one of the fissile materials. The remainder of the fuel must have a high mass number in order to avoid slowing down the neutrons. The natural reserves of fissionable materials are more than 100 times greater than the reserves of fissile materials. Consequently, from the viewpoint of utilization of available energy resources, fast reactors are of great importance. Breeder reactors are described later.

Moderators. The most important slow-down mechanism is elastic scattering on elements of low mass number. Materials like light and heavy water, beryllium oxide, and graphite are used to slow down, or *thermalize* the neutrons to an energy of about 0.025 eV. As neutrons collide with the nuclei of these atoms, their kinetic energy and speed are gradually reduced until thermal equilibrium is achieved with the reactor structure. The fewer such collisions before deceleration is complete, the less chance of ^{238}U atoms absorbing neutrons.

Critical Mass. Thermal neutrons, which move like atoms in a low-pressure gas, diffuse throughout the reactor. They must be absorbed by a nucleus of the reactor structure, in which case they merely make that nucleus radioactive. Or, they may strike a fissionable atom of ^{235}U, causing fission and, in turn, releasing more neutrons to maintain the reaction. Should the number of neutrons absorbed by the moderator and ^{238}U be greater than about 1.5 excess neutrons emitted from each fission, the chain reaction will not be maintained. Therefore, the reactor core must be designed so that the mass of fuel will be just sufficient to ensure one neutron from each fission causing fission in another atom. A mass and configuration of fissionable material in which this occurs is termed the *critical mass*—or a reactor in which this condition is achieved is said to have "gone critical."

To measure a chain reaction, a multiplication factor k is used to indicate the ratio of neutrons in one generation to those in the preceding generation. Thus, in a constant chain reaction where the total number of neutrons neither increases nor decreases, the heat output is constant and $k = 1$. Should k rise above unity, the rate of fission, and hence the rate of heat productivity, steadily rises. This is so even if k is held constant at its new value. Here lies one *major difference between nuclear reactors and conventional steam generators*. In the latter, heat output is proportional to firing rate. If the firing rate is increased, the steam output is increased; but it remains constant at its new level. In a nuclear reactor, an increase in k results in continuously rising heat output. Only by returning the rate of neutron production to its original ratio can heat output be maintained at its new level.

Reactivity Control. Absorption of excess neutrons, above those needed to maintain a constant reactivity level, provides close control over the degree of reactivity. This is accomplished by inserting materials having a high neutron-capture rate into the core. Control rods of special alloy metals are moved into and out of the cores as required. To start the reactor from shutdown (black start), control rods are partially withdrawn until *k* becomes greater than one. Neutron flux and heat output grow until the desired level is reached. At this point, control-rod movement is quickly reversed to keep *k* at unity. The reactor is shut down by inserting the rods to their full extent. In this position, the rods absorb more than 1.5 excess neutrons per fission and the chain reaction quickly stops. Heat production continues for a time, but is usually dissipated by an auxiliary cooling system.

It is interesting to contrast a nuclear power reactor and a nuclear bomb. The designer of a nuclear fission bomb seeks to release as much fission energy as possible within the shortest possible time (milliseconds). Thus, a bomb is designed to favor *prompt neutrons*. By contrast, the normal operating mode of a nuclear power reactor is one in which prompt neutrons alone cannot sustain a chain reaction, but prompt neutrons together with delayed neutrons can. Only the delayed neutrons are controllable. A power reactor is designed to release fission energy *slowly and smoothly* and in just the right amounts to convert water into steam. Whereas the "fuel" in a bomb is used up essentially in an instant, in a power reactor the energy release is spread over months and years. It has been agreed by physicists for many years that it is physically impossible for a power reactor to explode in the manner of an atomic bomb.

TYPES AND MAJOR CHARACTERISTICS OF NUCLEAR POWER REACTORS

In order, the following types of nuclear fission reactors are described in this section: (1) light water reactors, (a) pressurized water reactors, (b) boiling-water reactors; (2) high-temperature gas-cooled reactors; (3) heavy water reactors; and (4) fast breeder reactors. Military reactors are not described.

CONTEMPORARY LIGHT-WATER REACTORS (LWRs)[1]

These reactors are of two principal designs: (1) *pressurized water reactors* (PWR), and (2) *boiling-water reactors* (BWR). In a PWR, heat generated in the nuclear core is removed by water (reactor coolant) circulating at high pressure through the primary circuit. The water in the primary circuit both cools and moderates the reactor. Heat is transferred from the primary to the secondary system in a heat exchanger, or boiler, thereby generating steam in the secondary system. The BWR differs from the PWR primarily in that boiling takes place in the reactor itself. Comparable steam temperatures are possible at pressures of about 1000 pounds per square inch (6.9 mPa) as contrasted with 2000 psi (13.8 mPa) for pressurized reactors.

[1]In nuclear power technology, ordinary water, in contrast with *heavy water*, is termed *light water*.

Contemporary Boiling Water Reactor (BWR)

Aside from its heat source, the boiling water reactor (BWR) generation cycle is substantially similar to that found in fossil-fueled power plants. One of the first BWRs was the *Vallecitos* BWR, a 1000 psi (6.9 mPa) reactor which powered a 5 MW electric generator and provided power to the Pacific Gas & Electric Company grid through 1963. Power output capabilities have increased many times during the intervening years as shown by tabular summaries given later in this entry.

The direct-cycle boiling water reactor nuclear system (Fig. 1) is a steam generating system consisting of a nuclear core and an internal structure assembled within a pressure vessel, auxiliary systems to accommodate the operational and safeguard requirements of the nuclear reactor, and necessary controls and instrumentation. Water is circulated through the reactor core producing saturated steam which is separated from the recirculation water, dried in the top of the vessel, and directed to the steam turbine-generator. The turbine employs a conventional regenerative cycle with condenser deaeration and condensate demineralization. The direct-cycle system is used because of its inherently simple design, contributing to reliability and availability.

The steam from a BWR is, of course, radioactive. The radioactivity is primarily ^{16}N, a very short-lived nitrogen isotope (7 seconds half-life) so that the radioactivity of the steam system exists only during power generation. Extensive generating experience has demonstrated that shutdown maintenance on a BWR turbine, condensate, and feedwater components can be performed essentially as a fossil-fuel plant.

The reactor core, the source of nuclear heat, consists of fuel assemblies and control rods contained within the reactor vessel and cooled by the recirculating water system. A 1,220-MWe BWR/6 core consists of 732 fuel assemblies and 177 control rods, forming a core array 16 feet (4.8 meters) in diameter and 14 feet (4.2 meters) high. The power level is maintained or adjusted by positioning control rods up and down within the core. The BWR core power level is further adjustable by changing the recirculation flow rate without changing control rod position, a feature which contributes to excellent load-following capability.

The BWR is the only light water reactor system that employs bottom-entry control rods. From the very first BWRs, bottom-entry control

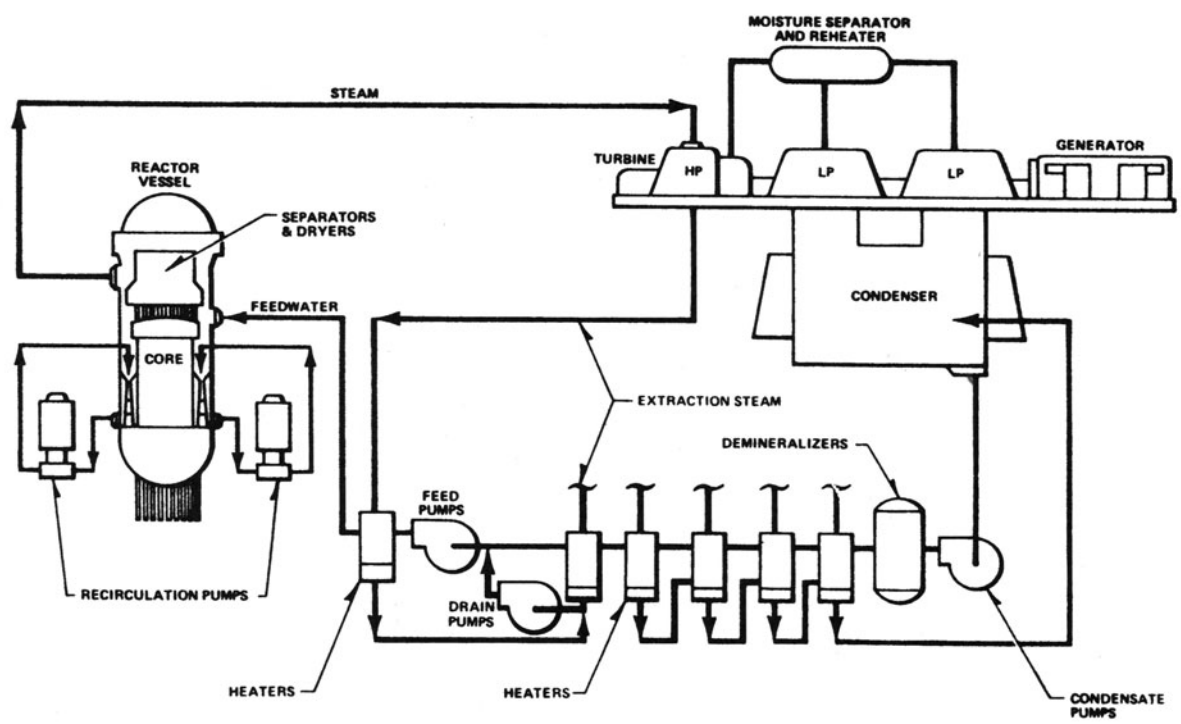


Fig. 1. Contemporary direct-cycle reactor system used in boiling water reactor. (*General Electric*.)

rods have been used because reactivity and moderator density is highest in the lower part of the core. They provide optimum power shaping characteristics for the type of core where moderator density is varied as a function of power level. Bottom-entry and bottom-mounted control rod drives also allow refueling without removal of rods and drives, and allow drive testing with an open vessel prior to initial fuel loading, or at each refueling operation. The hydraulic system, using reactor system pressure, provides rod insertion forces that are greater than gravity or mechanical systems.

The BWR requires substantially lower primary coolant flow through the core than pressurized water reactors. The core flow of a BWR is the sum of the feedwater flow and the recirculation flow, which is typical of any boiler. Unique to the BWR is the application of jet pumps inside the reactor vessel. See Fig. 2. The jet pumps deliver their driving force from the external recirculation pumps and generate about two-thirds of the recirculation flow within the reactor vessel. The jet pumps also contribute to the inherent safety of the BWR design under loss-of-coolant emergency conditions because they continue to provide internal circulation with one or both external recirculation loops out of service. The BWR can deliver about one-third power through this natural jet pump circulation mode, a vital capability in effecting a "black start" (a fully fresh start-up of a reactor) of the plant without external power.

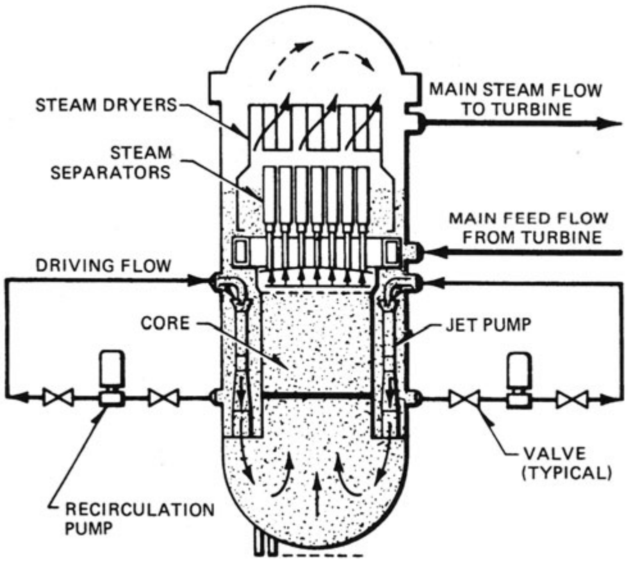


Fig. 2. Steam and recirculation water flow paths of contemporary boiling water reactor. (*General Electric*.)

The BWR operates at constant pressure and maintains constant steam pressure similar to most fossil-fueled boilers. The BWR primary system operates at pressure about one-half that of a pressurized water reactor primary system, while producing steam of equal pressure and quality.

The integration of the turbine pressure regulator and control system with the reactor water recirculation flow control system permits automated changes in steam flow to accommodate varying load demands on the turbine. Power changes of up to 25% can be accomplished automatically by recirculation flow control alone, at rate of 15% per minute increasing and 60% per minute decreasing. This provides a load-following capability that can track rapid changes in power demand.

Nuclear Boiler Assembly. This assembly consists of the equipment and instrumentation necessary to produce, contain, and control the steam required by the turbine-generator. The principal components of the nuclear boiler are: (1) reactor vessel and internals—reactor pressure vessel, jet pumps for reactor water circulation, steam separators and dryers, and core support structure; (2) reactor water recirculation system— pumps, valves, and piping used in providing and controlling core flow; (3) main steam lines—main steam safety and relief valves, piping, and pipe supports from reactor pressure vessel up to and including the isolation valves outside of the primary containment barrier; (4) control rod drive system—control rods, control rod drive mechanisms and hydraulic system for insertion and withdrawal of the control rods; and (5) nuclear fuel and in-core instrumentation.

Reactor Assembly. This assembly (Fig. 3) consists of the reactor vessel, its internal components of the core, shroud, top guide assembly, core plate assembly, steam separator and dryer assemblies and jet pumps. Also included in the reactor assembly are the control rods, control rod drive housings and the control rod drives.

Each fuel assembly that makes up the core rests on an orificed fuel support mounted on top of the control rod guide tubes. Each guide tube, with its fuel support piece, bears the weight of four assemblies and is supported by a control rod drive penetration nozzle in the bottom head of the reactor vessel. The core plate provides lateral guidance at the top of each control rod guide tube. The top guide provides lateral support for the top of each fuel assembly.

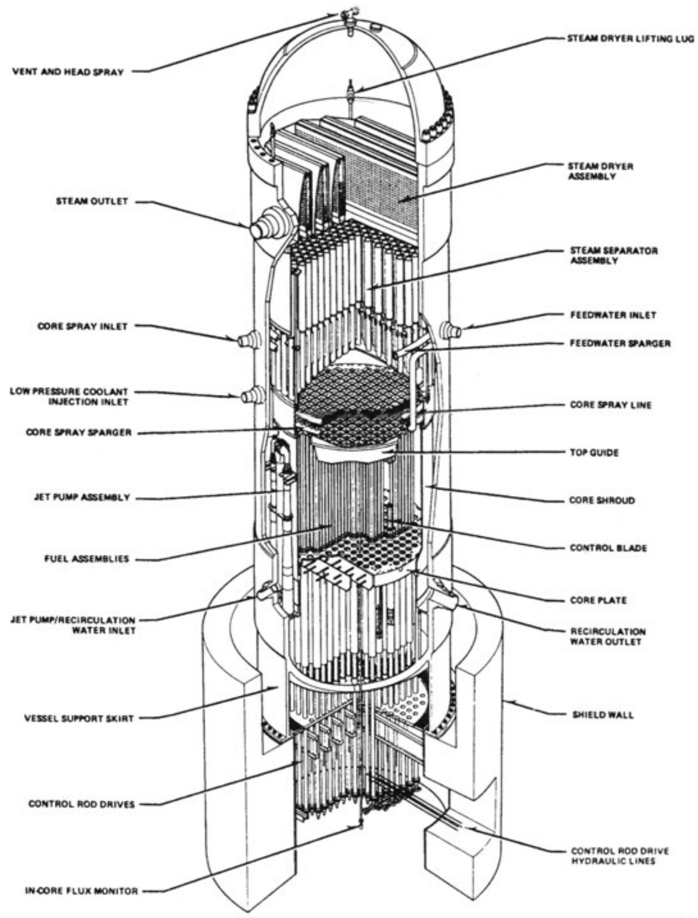


Fig. 3. Reactor assembly of contemporary boiling water reactor. (*General Electric*.)

Control rods occupy alternate spaces between fuel assemblies and may be withdrawn into the guide tubes below the core during plant operation. The rods are coupled to control rod drives mounted within housings which are welded to the bottom head of the reactor vessel. The bottom-entry drives do not interfere with refueling operations. A flanged joint is provided at the bottom of each housing for ease of removal and maintenance of the rod drive assembly.

Except for the Zircaloy in the reactor core, the reactor internals are stainless steel or other common corrosion-resistant alloys. The reactor vessel is a pressure vessel with a single full-diameter removable head. The base material of the vessel is low alloy steel which is clad on the interior except for nozzles with stainless steel weld overlay to provide the necessary resistance to corrosion.

The shroud is a cylindrical, stainless steel structure which surrounds the core and provides a barrier to separate the upward flow through the core from the downward flow in the annulus. Two ring spargers, one for low pressure core sprays and the other for high pressure core spray are mounted inside the core shroud in the space between the top of the core and steam separator base. The core spray ring spargers are provided with spray nozzles for the injection of cooling water under emergency conditions. A nozzle for the emergency injection of neutron absorber (sodium pentaborate) solution is mounted below the core in the region of the recirculation inlet plenum.

The steam separator assembly consists of a domed base on top of which is welded an array of standpipes with a 3-stage separator located at the top of each standpipe. The steam separator assembly rests on the top flange of the core shroud and forms the cover of the core discharge plenum region. In each separator, the steam-water mixture rising through the standpipe impinges on vanes which give the mixture a spin to establish a vortex wherein the centrifugal forces separate the water from the steam in each of three stages. Steam leaves the separator at the top and passes into the wet steam plenum below the dryer. The separated water exits from the lower end of each stage of the separator and enters the pool that surrounds the standpipes to join the downcomer annulus flow.

The steam dryer assembly is mounted in the reactor vessel above the separator assembly and forms the top and sides of the wet steam plenum. Vertical guides on the inside of the vessel provide alignment for the dryer assembly during installation. The dryer assembly is supported by pads extending inward from the vessel wall and is held down in position during operation by the vessel head. These vanes are attached to a top and bottom supporting member forming a rigid, integral unit. Moisture is removed and carried by a system of troughs and drains to the pool surrounding the separators and then into the recirculation downcomer annulus between the core shroud and reactor vessel wall.

Control Rod Drive System. Positive core reactivity control is maintained by the use of movable control rods interspersed throughout the core. These control rods thus control the overall reactor power level and provide the principal means of quickly and safely shutting down the reactor. The rods are vertically moved by hydraulically actuated, locking piston type drive mechanisms. The drive mechanisms perform both a positioning and latching function, and a scram function with the latter overriding any other signal (scram signifies prompt shutdown).

Core Configuration. The reactor core of the BWR is arranged as an upright cylinder containing a large number of fuel assemblies and located within the reactor vessel. The coolant flows upward through the core. The plan of a typical core arrangment of a large BWR is shown in Fig. 4. The lattice configuration is shown in Fig. 5.

Fuel Rod. A fuel rod consists of uranium dioxide (UO_2) pellets and a

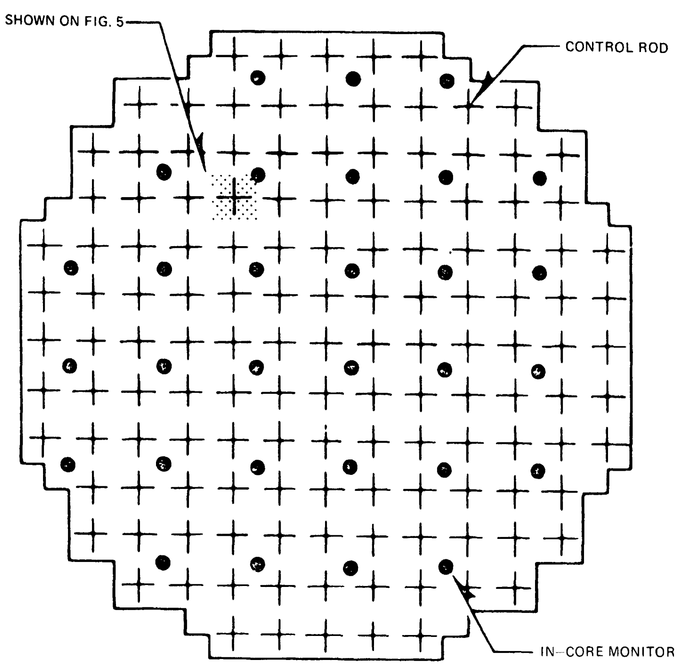


Fig. 4. Typical core arrangement in contemporary boiling water reactor. (*General Electric.*)

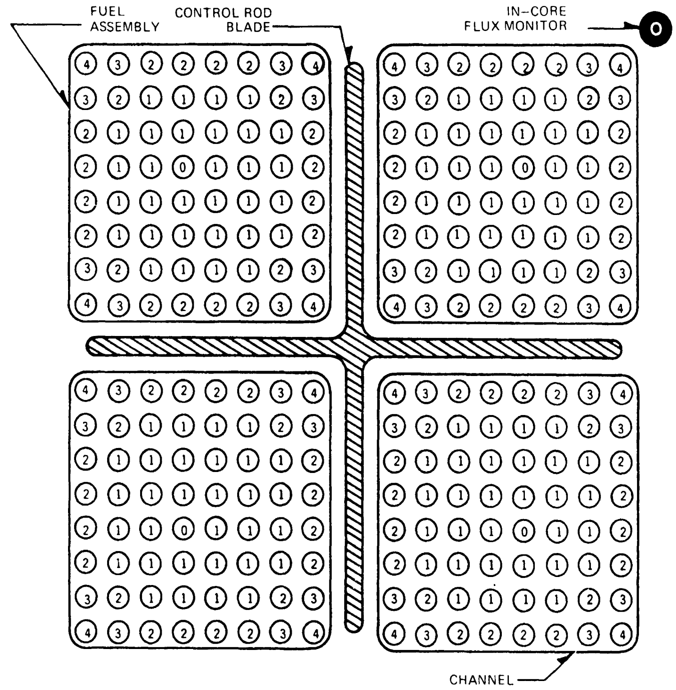


Fig. 5. Core lattice arrangement in contemporary boiling water reactor. (*General Electric.*)

Zircaloy-2 cladding tube. The UO_2 pellets are manufactured by compacting and sintering UO_2 powder into cylindrical pellets and grinding to size. The immersion density of the pellets is approximately 95% of theoretical UO_2 density. A fuel rod is made by stacking pellets into a cladding tube which is evacuated, back-filled with helium to atmospheric pressure, and sealed by welding Zircaloy end plugs in each end of the tube. The pellets are stacked to an active height of 148 inches (376 centimeters) with the top 12 inches (30.5 centimeters) of tube available as a fission gas plenum. A plenum spring is provided in the plenum space to exert a downward force on the pellets, the spring keeping the pellets in place during the preirradiation handling of the fuel bundle.

Fuel Bundle. Each fuel bundle contains 63 fuel rods which are spaced and supported in a square (8×8) array by a lower and upper tie plate. Three types of rods are used in a fuel bundle: (1) tie rods; (2) a water rod; and (3) standard fuel rods. The third and sixth fuel rods along each outer edge of a bundle are tie rods. The eight tie rods in each bundle have threaded-end plugs which screw into the lower tie plate casting. A stainless steel hexagonal nut and locking tab is installed on the upper end plug to hold the assembly together. The water rod not only serves as a spacer support rod, but also provides a source of moderator material near the center of the fuel bundle. This flattens the neutron flux across the bundle, and leads to lower local peaking factors and better utilization of uranium in the interior rods of the fuel assembly.

The initial core will contain fuel assemblies having a common average enrichment ranging from approximately 1.6% (weight) of ^{235}U to 2.2%, depending upon initial cycle requirements. Each assembly will contain different enrichment rods. Selected rods in each assembly will, in addition, be blended with gadolinium burnable poison. The reload fuel will also contain four different enrichment rods with an average enrichment in the range of 2.4 to 2.8%. Different ^{235}U enrichments are used in fuel assemblies to reduce the local power peaking. Low enrichment rods are used in the corner rods and in the rods nearer the water gaps; higher enrichment uranium is used in the central part of the fuel bundle.

Fuel Channel. A fuel channel encloses the fuel bundle. The combination of a fuel bundle and a fuel channel is called a fuel assembly. See Fig. 6. The channel is a square-shaped tube fabricated from Zircaloy 4. The outer dimensions are 5.518 inches (14 centimeters) by 5.518 inches (14 centimeters) by 166.9 inches (424 centimeters) long. The reusable channel makes a sliding seal fit on the lower tie plate surface. It is attached to the upper tie plate by the channel fastener assembly, consisting of a spring and a guide, and a cap screw secured by a lock washer.

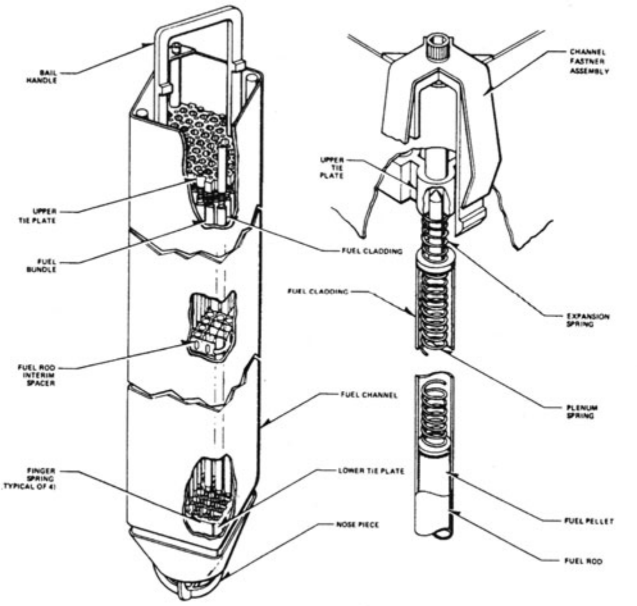


Fig. 6. Fuel assembly of contemporary boiling water reactor. (*General Electric.*)

The fuel channels direct the core coolant flow through each fuel bundle and also serve to guide the control rods.

Neutron Sources. Several antimony-beryllium start-up sources are located within the core. They are positioned vertically in the reactor by "fit up" in a slot (or pin) in the upper grid and a hole in the lower core support plate. The active portion of each source consists of a beryllium sleeve enclosing two antimony-gamma sources. The resulting neutron emission strength is sufficient to provide indication on the source range neutron detectors for all reactivity conditions equivalent to the condition of all rods inserted prior to initial operation. The active source material is entirely enclosed in a stainless steel cladding with an outside diameter of approximately 0.7 inch (1.8 centimeter). The source is cooled by natural circulation of the core leakage flow in the annulus between the beryllium sleeve and the antimony-gamma sources.

Core Design Margins. The reactor core is designed to operate at rated power with sufficient design margin to accommodate changes in reactor operations and reactor transients without damage to the core. In order to accomplish this objective, the core is designed, under the most limiting operating conditions and at 100% rated power, to meet the following bases: (1) The maximum linear heat generation rate, in any part of the core, is always less than 13.4 kW/foot (43.97 kW/meter); and (2) the minimum ratio between critical heat flux and fuel operating heat flux, in any part of the core, is always greater than 1.9.

Power Distribution. The design power distribution is divided for convenience into several components: (1) relative assembly power; (2) local; and (3) axial. The relative assembly power peaking factor is the maximum fuel assembly average power divided by the reactor core average assembly power. The local power peaking factor is the maximum fuel rod average heat flux in an assembly divided by the assembly average fuel rod heat influx. The axial power peaking factor is the maximum heat flux of a fuel rod divided by the average heat flux in that rod. Peaking factors vary throughout an operating cycle, even at steady-state full-power operation, since they are affected by withdrawal of control rods to compensate for fuel burnup. The design peaking factors represent the values of the most limiting power distribution that will exist in the core throughout its life.

Because of the presence of steam voids in the upper part of the core, there is a natural characteristic for a BWR to have the axial power peak in the lower part of the core. During the early part of an operating cycle, bottom-entry control rods permit a partial reduction of this axial peaking by locating a larger fraction of the control rods in the lower part of the core. At the end of an operating cycle, the higher accumulated exposure and greater depletion of the fuel in the lower part of the core reduces the axial peaking. The operating procedure is to locate control rods so that the reactor operates with approximately the same axial power shape throughout an operating cycle.

Reactivity Control. The movable boron-carbide control rods are sufficient to provide reactivity control from the cold shutdown condition to the full-load condition. Supplementary reactivity control in the form of solid burnable poison is used only to provide reactivity compensation for fuel burnup or depletion effects. The movable control rod system is capable of bringing the reactor to the subcritical when the reactor is an ambient temperature (cold), zero power, zero xenon, and with the strongest control rod fully withdrawn from the core. In order to provide greater assurance that this condition can be met in the operating reactor, the core is designed to obtain a reactivity of less than 0.99, or a 1% margin on the "stuck rod" condition. See Fig. 7.

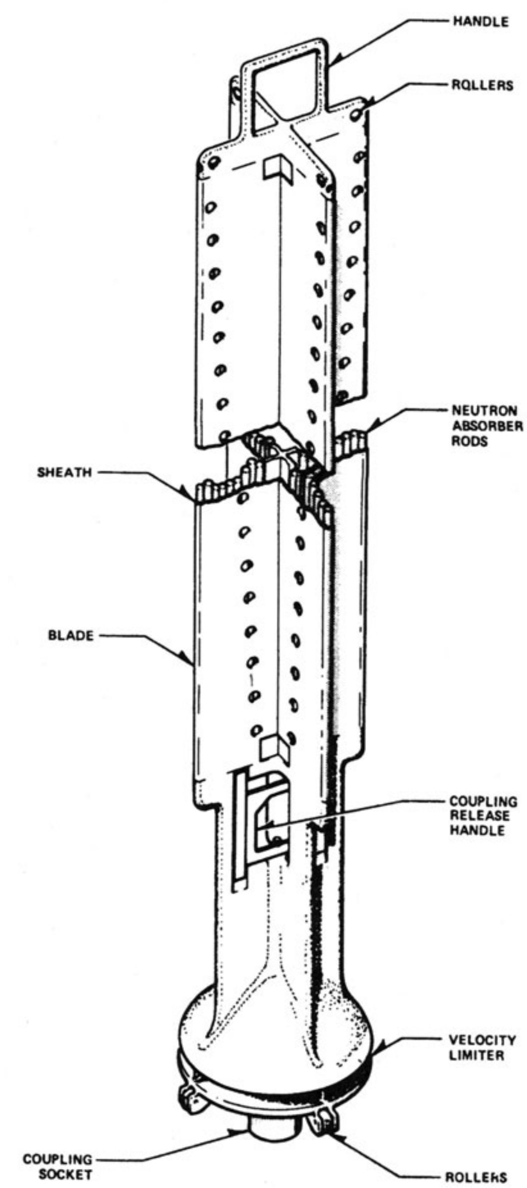


Fig. 7. Control rod used in contemporary boiling water reactor. The cruciform control rods contain 76 stainless steel tubes (19 tubes in each wing of the cross). These tubes are filled with boron carbide powder compacted to approximately 65% of theoretical density. The tubes are seal-welded with end plugs on either end. The individual tubes act as pressure vessels to contain the helium gas released by the boron-neutron capture reaction. The control rods have an active length of 144 inches (365.8 centimeters) of boron carbide, a span of 9.75 inches (24.8 centimeters), and an overall length of 173.75 inches (441.3 centimeters). The control rods can be positioned at 6-inch (15-centimeter) steps and have a nominal withdrawal and insertion speed of 3 inches (7.5 centimeters) per second. Control rods are cooled by the core leakage (bypass) flow. In addition to satisfying initial control effectiveness requirements, it is expected that the control rods will have an average lifetime of approximately 15 full-power years. (*General Electric.*)

Reactor Auxiliary Systems include (1) a reactor water cleanup system for maintaining high reactor water quality by removing fission products, corrosion products, and other soluble and insoluble impurities; (2) a fuel and containment pool cooling and cleanup system—a system which accommodates the beta and gamma radiation heating from the fission products that remain in the spent fuel, as well as drywell heat transferred to the upper containment pool; (3) a closed cooling water system for reactor service consisting of a separate, force circulation loop; (4) emergency equipment cooling system; (5) standby liquid control system; (6) reactor core isolation cooling system; (7) emergency core cooling system; (8) high-pressure core spray system; and (9) residual heat removal system.

Contemporary Pressurized Water Reactor (PWR)

In a typical pressurized water reactor (PWR), heat generated in the nuclear core is removed by water (reactor coolant) circulating at high pressure through the primary circuit. The water in the primary circuit cools and moderates the reactor. The heat is transferred from the primary to the secondary system in a heat exchanger, or boiler, thereby generating steam in the secondary system. The steam produced in the steam generator, a tube-and-shell heat exchanger, is at a lower pressure and temperature than the primary coolant. Therefore, the secondary portion of the cycle is similar to that of the moderate-pressure fossil-fueled plant. In contrast, in boiling-water or direct-cycle systems, steam is generated in the core and is delivered directly to the steam turbine.

The similarities of basic pressurized water reactor design from one manufacturer to the next are more striking than the differences. Therefore, the description of one particular configuration (Combustion Engineering, Inc.) can suffice to convey the general operating principles. The major components of a PWR are: (1) the reactor vessel which contains the oxide fuel core, core intervals, control element assemblies, and in-core instruments; (2) the electrically-heated pressurizer; (3) the electric-motor-driven primary coolant pumps; and (4) the U-tube type steam generators. See Fig. 8. The primary coolant system layout can be fitted into a variety of containment types and concepts. A prestressed cylindrical containment is common. Figure 9 shows the arrangement in a spherical containment. This type of building lends itself to separation of safeguards equipment, steam lines, and emergency power supplies.

Steam Generators. The basic geometry is shown in Fig. 10. With the nuclear steam supply system operating at 3,817 MW, two steam generators produce a total of 17.18 × 10^6 pounds (7.89 × 10^6 kilograms) of steam per hour at 1,070 psia (72.8 atmospheres). The steam generators are constructed, using carbon steel pressure-containing members and Inconel-600 tubes. The tube-sheet is clad by weld deposit for maximum strength; tongue and groove construction of the divider plate places no stress on the tube-sheet cladding. Fusion welding of the end of each tube to the tube-sheet primary cladding provides an effective seal for leakage control, and "expanding" (explosively expanding) the tubes in the full length of the tube-sheet eliminates corrosion-prone crevices. An economizer section on the units improves heat transfer by preheating the incoming feedwater, using the low (primary side) temperature heat transfer area of the U-tubes. Multiple feed nozzles allow the economizer flow distribution to be optimized for each power level.

Reactor Coolant Pumps. As indicated by Fig. 8, four reactor coolant pumps are used, two for each steam generator. The pumps are vertical, single-bottom-suction, horizontal-discharge, motor-driven centrifugal units. The pump impeller is keyed and locked to its shaft. A complex system of seals is used to prevent any leakage. The motors are designed to start and accelerate to speed under full load with a drop to

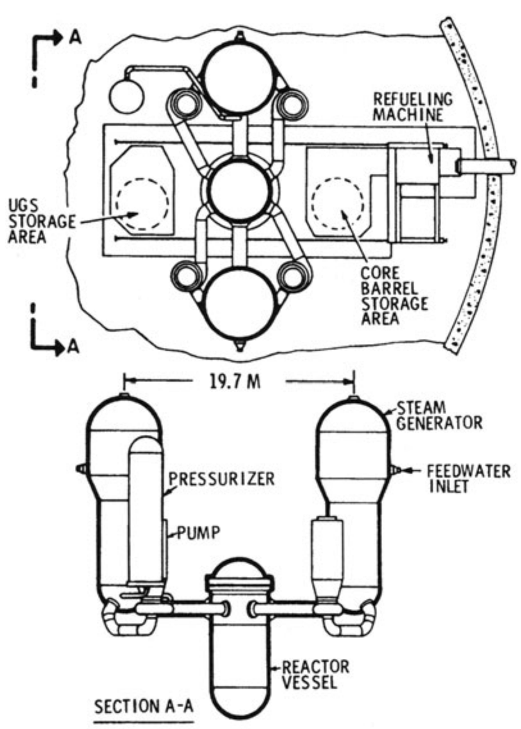


Fig. 8. Nuclear steam supply system for contemporary pressurized water reactor. (*Combustion Engineering.*)

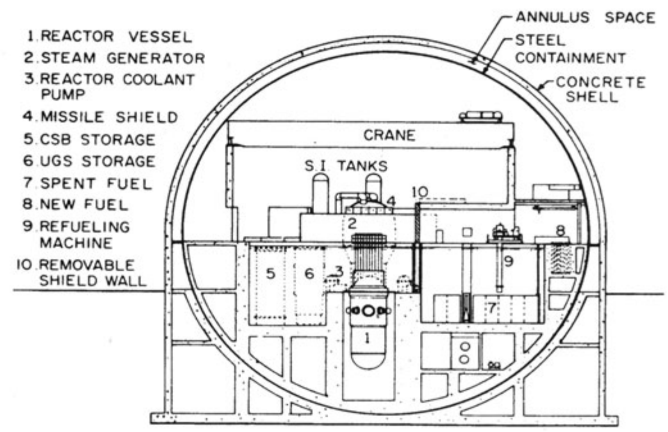


Fig. 9. Spherical containment for contemporary pressurized water reactor. (*Combustion Engineering.*)

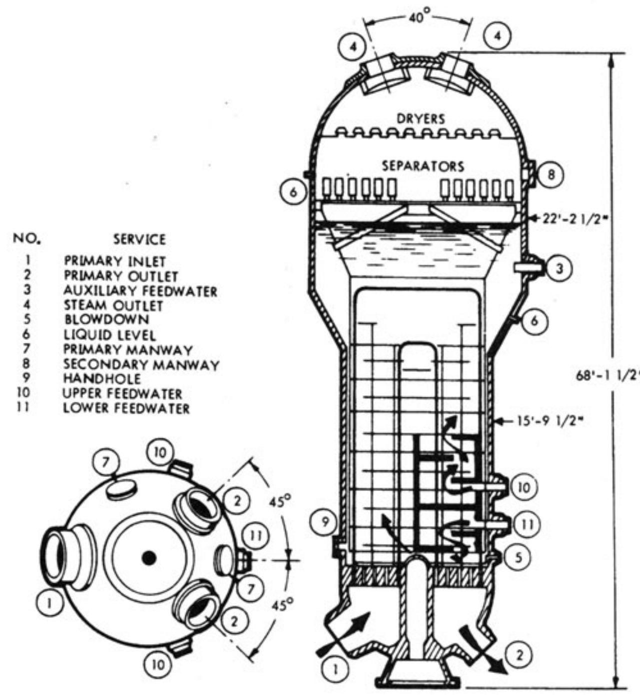


Fig. 10. Steam generator for contemporary pressurized water reactor. (*Combustion Engineering.*)

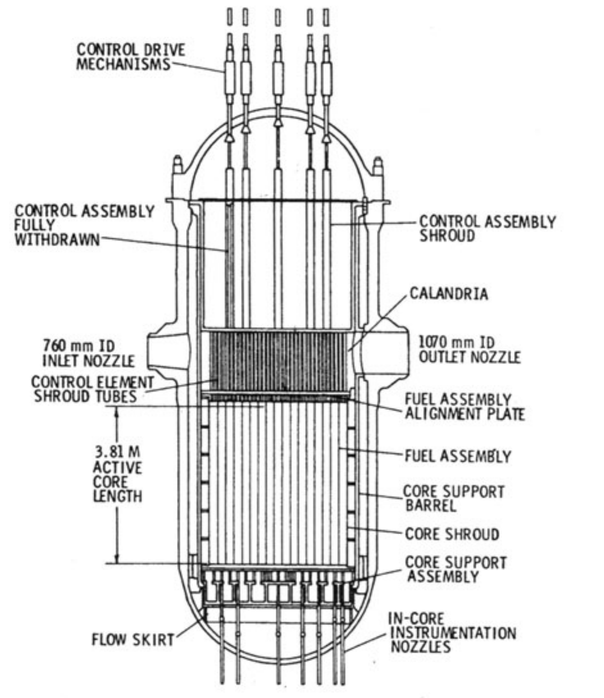


Fig. 11. Reactor arrangement in contemporary pressurized water reactor.

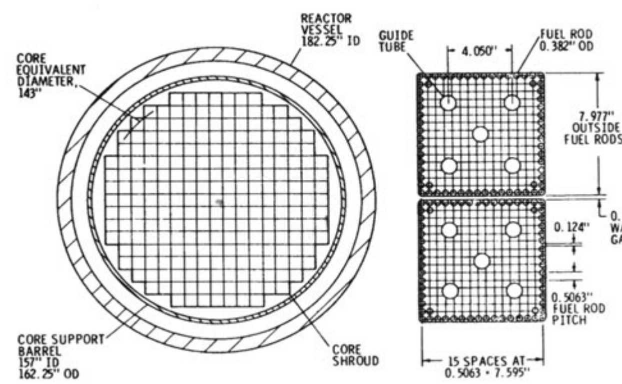


Fig. 12. Reactor core cross section of contemporary pressurized water reactor with 241 fuel assemblies. (*Combustion Engineering.*)

80% of normal rated voltage at the motor terminals. Each motor is provided with an anti-reverse rotation device. Each reactor coolant pump is provided with four vertical support columns, four horizontal support columns, and one vertical snubber. The structural columns provide support for the pumps during normal operation, earthquake conditions, and any hypothetical loss-of-coolant accident in either the pump suction or discharge line.

Pressurizer. The pressurizer is a cylindrical pressure vessel, vertically mounted and bottom supported. Energy to the water is supplied by replaceable direct-immersion electric heaters which are inserted from the bottom head of the pressurizer. Nozzles are provided for spray, surge, relief, and instrumentation connections. The pressurizer maintains reactor coolant system operating pressure and, in conjunction with the chemical and volume control system, compensates for changes in reactor coolant volume during load changes, heat-up, and cool-down. During full-power operation, the pressurizer is about $\frac{1}{3}$ full of saturated steam.

Reactor Vessel. This vessel is designed to contain the fuel bundles, the control element assembles, and the internal structures necessary for support of the core. The reactor is a stainless clad, thick-walled, carbon steel pressure vessel comprised of a cylinder with two hemispherical heads. The lower head is integrally welded to the vessel shell and contains in-core instrumentation nozzles; the upper closure head, containing the control element drive mechanism nozzles, is attached to the vessel by means of a bolted flange, thus permitting the head to be removed to provide access to the reactor internals. The head flange is drilled to match the vessel flange stud bolt locations.

The vessel flange is a forged ring with a machined ledge on the inside surface to support the core support barrel. The flange is drilled and tapped to receive the closure studs and is machined to provide a mating surface for the reactor vessel closure seals. Sealing is accomplished by using two silver-plated, NiCrFe-alloy, self-energizing O-rings. The space between the two rings is monitored to detect any inner-ring coolant leakage. The inlet and outlet nozzles are located radially on a common plane below the vessel flange. Extra thickness in the vessel course provides the reinforcement required for the nozzles. Snubbers built into the lower portion of the vessel shell limit the amplitude of any displacement of the core support barrel. Core stops are also built into the reactor vessel to limit the downward displacement of the core support barrel.

Cladding for the reactor vessel is a continuous integral surface of corrosion-resistant material, having $\frac{7}{32}$-inch (0.56 centimeter) nominal thickness, and a $\frac{1}{8}$-inch minimum thickness. The reactor vessel is supported by four vertical columns located under the vessel inlet nozzles. These columns are designed to flex in the direction of horizontal thermal expansion and thus allow unrestrained heat-up and cool-down. The columns also act as a hold-down device for the vessel. The supports are designed to accept normal loads and seismic and pipe rupture accident loads.

The reactor arrangement is shown in Fig. 11. The barrel-calandria guide structure is a rugged (3-inches thick, barrel section) unit which can withstand and protect all control element fingers from the combined effects of seismic and blowdown loads that may result from a loss-of-cooling accident. The calandria structure fits over the control element guide tubes of the fuel assemblies, aligning all fuel assemblies, and laterally restraining the top ends of the fuel assemblies. With the upper guide structure in place, a continuous guide tube for each control finger is formed, extending from the top of the tube-sheet to the bottom of the fuel assembly. Because of this feature, which isolates every control finger from the coolant crossflow, flexibility is obtained in the number of control fingers that can be attached to one control assembly, i.e., one control element assembly can serve more than one fuel assembly.

Severe emergency core cooling system criteria require that the builders of water reactors increase the linear feet of fuel in the reactor core for the same power in order to reduce LOCA (loss-of-cooling accident) fuel temperatures. In the unit described here, an assembly with a 16 × 16 fuel rod array of smaller diameter rods is used in the same assembly envelope that was occupied by a 14 × 14 assembly in earlier designs. This results in a maximum linear heat rate decrease in the assembly of about 25%.

As shown in Fig. 12, the active core is made up of 241 fuel assemblies all of which are mechanically identical. As indicated by Fig. 13, each fuel assembly contains 236 Zircaloy-clad, UO_2 fuel rods retained in a structure consisting of Zircaloy spacer grids welded at about 15-inch (38.1-centimeter) intervals to five Zircaloy control element assembly guide tubes which, in turn, are mechanically fastened at each end to stainless steel end fittings. The overall length of the fuel assembly is about 177 inches (450 centimeters) and the cross section is about 8 inches (20.3 centimeters) by 8 inches (20.3 centimeters). Each fuel assembly weighs about 1,450 pounds (657.7 kilograms). With reference to Fig. 13, fuel rods, consisting of uranium dioxide (UO_2) pellets of low enrichment canned in thin-walled Zircaloy-4 tubing, are designed to achieve average burnups of about 33,000 MWD/MTU (thermal megawatt days/metric tons of uranium) and peak burnups of about 50,000 MWD/MTU. The design factors limiting burnup of the fuel are the effects on the clad of volumetric changes of the fuel pellet and fission gas release.

As indicated in Fig. 13, the fuel rod consists essentially of 0.325-inch (0.82-centimeter) diameter, 0.390-inch long UO_2 pellets canned in a 0.382-inch (0.97-centimeter) outside diameter Zircaloy-4 tube. The high density fuel pellets are dished at both ends to allow for axial differential thermal expansion and fuel volumetric growth with burnup.

The control element assemblies consist of an assembly of 4, 8, or 12

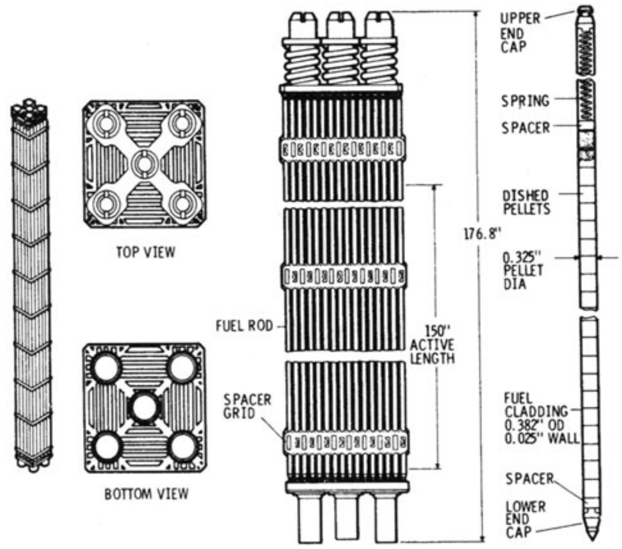


Fig. 13. Fuel assembly used in contemporary pressurized water reactor.

fingers approximately 0.8-inch (2-centimeter) outside diameter and arranged as shown in Fig. 14. The use of cruciform control rods, as in boiling water and early pressurized water reactors, necessitates large water gaps between the fuel assemblies to ensure that the control rods will scram (prompt shutdown) satisfactorily. These gaps cause peaking of the power in fuel rods adjacent to the water channel compared to fuel rods some distance from the channel.

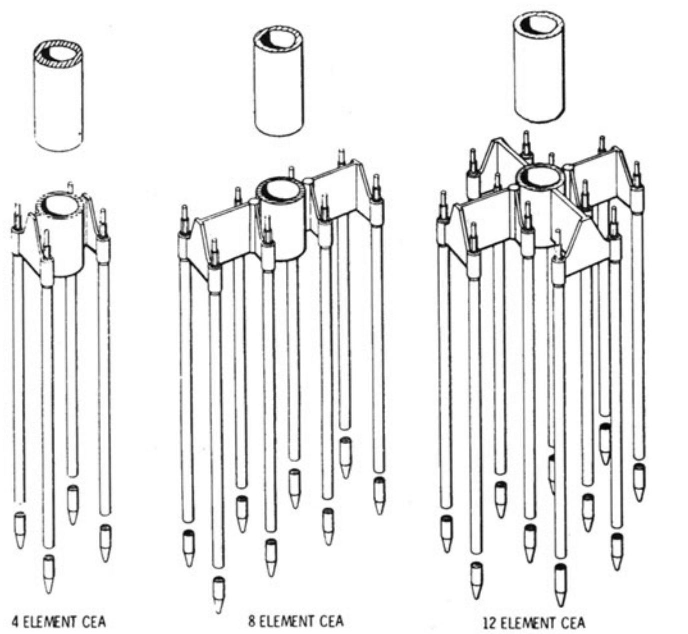


Fig. 14. Four 8- and 12-element control element assemblies. (*Combustion Engineering.*)

A five-hole assembly design was evolved from consideration of the lower peaking effect of smaller (removal of one fuel rod) water holes versus the mechanical advantages of larger (four fuel rods removed) water holes. The larger water holes allowed the use of rugged, 0.9-inch (2.3-centimeter) outside diameter by 0.035-inch (0.89-centimeter) thick Zircaloy guide tubes for the fuel assembly structure. These water holes are distributed relatively uniformly in the reactor core when placed in the 16 × 16 fuel rod lattice. The particular arrangement of water holes was selected in consideration of the water gap between fuel assemblies; the effect of the central water hole in the fuel assembly is balanced by the water gap between fuel assemblies. The mechanical simplicity and ruggedness outweigh the advantage of obtaining a small decrease in local peaking by using very small fingers. The slightly higher peaking associated with the design can be compensated, to a large extent, by varying the enrichment of the fuel in the rods adjacent to the control channel and/or by using water displacers in local hot spots.

The control element assembly, shown in Fig. 15, consists of 0.8-inch (2-centimeter) outside diameter Inconel tubes containing boron carbide pellets as the neutron absorbing material. A gas plenum is provided in order to limit the maximum stress due to generation of internal gas pressure. The individual control fingers are attached mechanically and locked to the various spider assemblies. This allows for simplifications in manufacture, shipping, and assembly of the control element assembly. Because all fingers are removable and replaceable, servicing and disposal problems are decreased. It is intended that the spider assembly and its extension shaft be reused whenever possible.

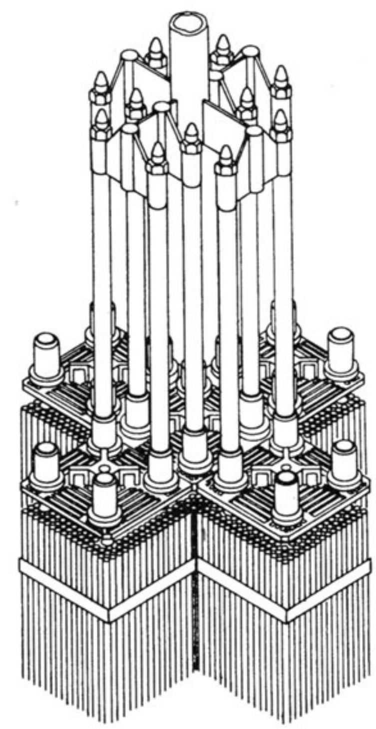


Fig. 15. Control element assembly and fuel for contemporary pressurized water reactor.

Design of the upper guide structure permits flexibility in the number of control fingers that can be attached to one control assembly. The standard pattern of control assemblies is shown in Fig. 16. In the standard design, power changes at close to full power, shaping of the radial power distribution, and control of the axial power distribution are best handled by the low worth 4-finger control element assembly entering a single fuel assembly. Shutdown reactivity control in the peipheral region of the core is handled by the 8-finger control element assembly and in the central region of the core by the 12-finger control element assembly. The need for the two types of shutdown control element assemblies is to obtain "stuck rod worths" in the high reactivity fuel on the periphery of the core which are about equal to the control element assembly control worth in the lower reactivity central zone of the core.

Instrumentation. The large size of present water reactors and the nuclear effects which can occur, such as xenon redistribution, stuck rods, and reactivity anomalies require that emphasis be placed on instrumentation and control systems if high plant availability is to be maintained, while providing the necessary protection due to abnormal occurrences. Because there are many reactivity effects that can produce changes in the reactor power distribution, more reliable operation can

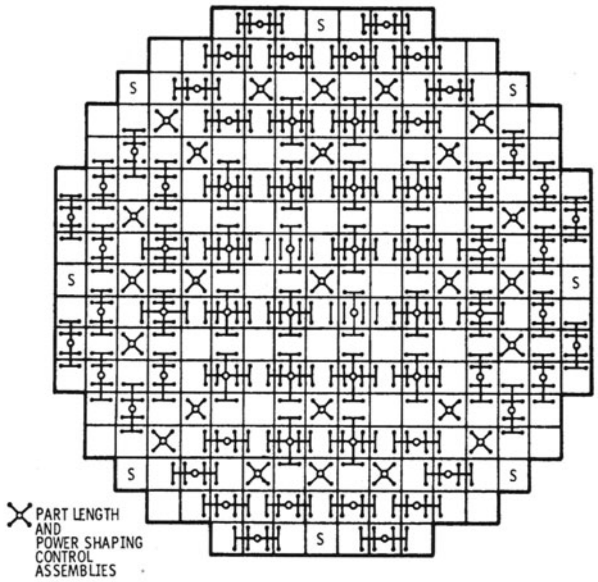


Fig. 16. Standard pattern of control assemblies in contemporary pressurized water reactor core. The pattern provides more-than-sufficient control for self-generated plutonium recycle. For complete open-market plutonium recycle, 4-element control assemblies are added in positions marked S.

be obtained by on-line monitoring of the reactor. This is best achieved with in-core instrumentation. The system described here has provision for up to 61 in-core instrument (ICI) assemblies which enter from the bottom of the reactor vessel. Radial distribution of the ICI is such that every type of fuel assembly, rodded and unrodded, is instrumented, assuming symmetrical core power distribution. Five sets of four symmetrically located ICI assemblies are included to monitor core power tilts. Also, every instrumented fuel assembly is either immediately adjacent to or diagonally adjacent to an instrumented fuel assembly to obtain good radial coverage of the core. Each of the ICI assemblies contains five self-powered fixed detectors distributed axially along the length of the core, a thermocouple at the end of the assembly to monitor outlet temperature, and a dry-well instrument tube which can accommodate a movable detector. This allows for high measurement accuracy of the on-line fixed detector.

The continuous monitoring and processing of the data from over 300 fixed detectors by the core monitoring computer provides the operator with information on core power distribution, maximum linear heat rate in the fuel, departure from nucleate boiling ratio, and fuel exposure. These data can then be used to obtain improved maneuvering of core power using the relative low worth 4-finger control element assemblies for power changes and power distribution control.

HIGH-TEMPERATURE GAS-COOLED REACTOR (HTGR)

Although there have been comparatively few gas-cooled reactors installed for generating commercial nuclear electric power, the concept has a number of operating advantages over light-water reactors and could play an important role in the reactor designs for the next century.

The high-temperature gas-cooled reactor (HTGR) is a thermal reactor that produces desired steam conditions. Helium is used as the coolant. Graphite, with its superior high-temperature properties, is used as the moderator and structural material. The fuel is a mixture of enriched uranium and thorium in the form of carbide particles clad with ceramic coatings.

The high-temperature conditions and high thermal efficiency (approximately 39%) of the (HTGR) result in high performance. The amount of cooling water required to carry away the waste heat is significantly less than in a light-water reactor (LWR). The use of thorium in the fuel cycle decreases fuel cost, improves the conservation of fuel, and adds the large deposits of thorium to available fuel reserves. The HTGR has significant environmental advantages, including: (1) lower thermal discharge because of its high efficiency; (2) low release of radioactive waste because of the high-integrity fuel and the inert coolant; and (3) low consumption of raw materials because of high efficiency and use of thorium in the fuel cycle.

High operating temperatures at moderate pressures are achieved through the use of helium as the coolant. Helium is attractive as a coolant because it: (1) is chemically inert; (2) absorbs essentially no neutrons; and (3) makes no contributions to the reactivity of the system. Carbon dioxide also has been used as a coolant.

Graphite is used as the moderator and core structural material because of (1) excellent mechanical strength at high temperatures; (2) very low neutron-capture cross section; (3) good thermal conductivity; and (4) high specific heat. Graphite has a long history of use in thermal reactors. Because of low neutron-capture cross section, no neutrons are lost within the core through absorption in metallic fuel cladding or structural supports. Graphite also is well suited to high-temperature operations, increasing in strength with temperature up to a point (2,482°C) well beyond the operating range of the HTGR.

The use of the thorium-uranium fuel cycle in the HTGR provides improved core performance over the plutonium/uranium low-enrichment cycle used in LWRs. The principal reason for this is that fissile ^{233}U produced from neutrons captured in thorium during reactor operation is neutronically a better fuel than ^{239}Pu, produced from ^{238}U in the low-enrichment cycle. The excellent neutronic characteristics of the graphite-moderated thorium/uranium cycle leads directly to high conversion ratios and low fuel inventories. Reduced ^{235}U inventories and make-up requirements spell reduced sensitivity to uranium prices.

Early Development of the HTGR

Work on the gas-cooled reactor has been underway essentially since the dawning of the nuclear power industry. The earliest developments were in Britain and France, at which time carbon dioxide gas was used as the coolant. In 1965, Britain opted for an advanced gas-cooled reactor (AGR). In 1969, France swung away from the HTGR (because of high construction costs) and targeted to the employment of more LWRs as well as commencing a concerted effort to develop a fast breeder type reactor. West Germany has been active in the development and testing of HTGRs since the early 1960s, but only recently (late 1980s) have the Germans indicated serious efforts toward commercialization of the HTGR. In the United States, a HTGR was installed at Peach Bottom, Pennsylvania, commencing commercial operation in 1974. As of the late 1980s, only one other HTGR was installed in the United States, the Fort St. Vrain plant near Denver, Colorado.

A simplified flow diagram of this station, which generates 842 MW (thermal) to achieve a net output of 330 MW (electrical), is given in Fig. 17. The helium coolant at a pressure of about 700 psi (47.6 atmospheres) flows downward through the reactor core where it is heated to 777°C. The coolant flow can be trimmed by the use of orifice valves located at the top of the core that are integral with the control rod drive mechanisms. From the reactor core, the coolant flows through the steam generators. After passing through the steam generators, the helium is returned to the core at a temperature of about 404°C by four steam-tur-

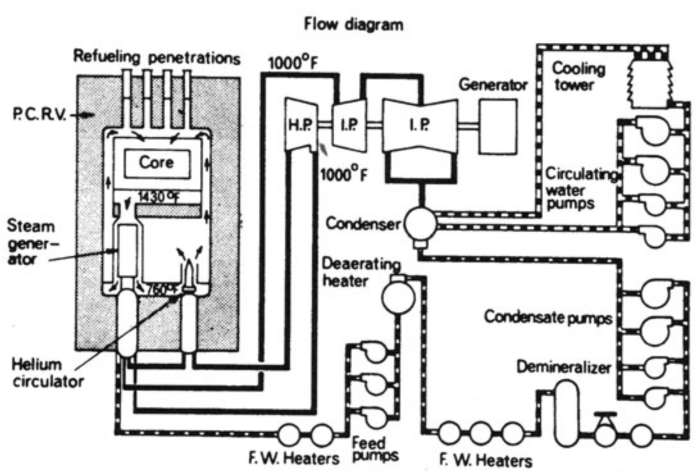


Fig. 17. Simplified flow diagram of Fort St. Vrain Nuclear Generating Station. (*GA Technologies*.)

bine-driven helium circulators. Two identical loops are used, each including a six-module steam generator and two helium circulators. Each loop contributes half of the total output of the nuclear steam supply system, which produces steam at 2,400 psig (163.3 atmospheres) and 538°C with single reheat to 538°C. The helium circulators are driven by the exhaust steam from the high-pressure turbine. This steam is then reheated and returned to the intermediate-pressure turbine. The circulators are also equipped with a Pelton water wheel drive so that they may be driven using the boiler feed pumps for emergency conditions.

The general reactor arrangement is shown in Fig. 18. The prestressed concrete reactor vessel (PCRV) is 31 feet in internal diameter with a 75-foot (23 meters) internal height. The upper and lower heads are nominally 15 feet (4.5 meters) thick, and the walls have a nominal thickness of 9 feet (2.7 meters). Thus, the PCRV provides the dual function of containing the coolant at operating pressure and also providing radiological shielding.

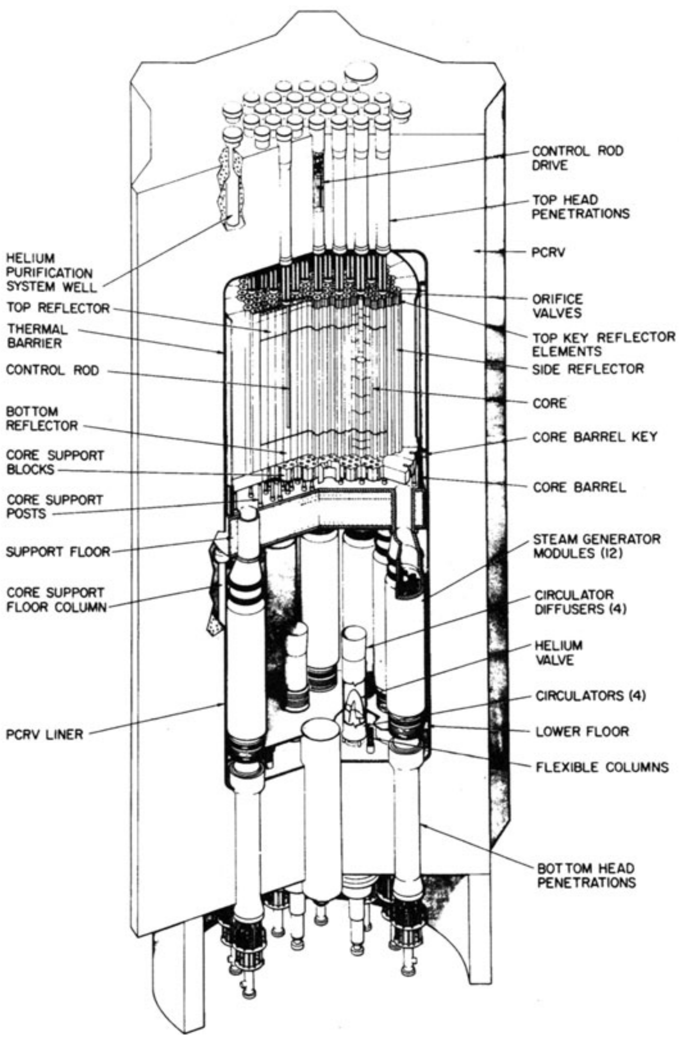


Fig. 18. General reactor arrangement of Fort St. Vrain high-temperature gas-cooled reactor. (*GA Technologies*.)

Reactor Core. The HTGR fuel element is a graphite block, hexagonal in cross section and having a grid of longitudinal fuel holes and coolant channels. The fuel element blocks are stacked in columns of eight blocks each and grouped into fuel regions consisting of a central column surrounded by six columns. Each region rests on a large core support block which, in turn, rests on graphite posts standing on the liner of the central cavity. Hexagonal graphite reflector elements are located above, below, and around the active core. These elements are surrounded by permanent side-reflector blocks to give the entire assembly a circular configuration. The fuel holes contain a rod consisting of ceramic-coated fuel particles in a graphite matrix. The coatings, applied by pyrolitic techniques, are multilayered to ensure a high degree of fission-product confinement. A porous interlayer, or buffer zone, accommodates the expansion of the irradiated fuel and provides storage space for gaseous fission products. The outer layer acts as a fission-product retention barrier and provides structural strength. In effect, the particle coating functions as a miniature spherical pressure vessel.

To achieve a fuel management scheme with the lowest fuel cycle cost consistent with the current thermal and material performance limits, the following parameters are selected: (1) a fuel cycle incorporating uranium/thorium; (2) a fuel lifetime of four years; (3) an average power density of 8.4 W/cm^3; and (4) a refueling frequency of once a year.

The reactor is controlled by two control rods located in each refueling region. All control rod pairs have scram (quick shutdown) capability and are driven by gravity. A backup reserve shutdown system is included. This consists of boronated graphite pellets that can be introduced from hoppers located in each refueling penetration into the core via the cylindrical channels in the central fuel element of each refueling region.

Safeguard Systems. The design of HTGR incorporates many inherent safety features and a number of engineered safeguards. The inherent safety characteristics include negative power and temperature coefficients, assured by the thorium content of the fuel. In addition, the high heat capacity of the large mass of graphite ensures that any core temperature transient resulting from reactivity insertions or interruptions in cooling will be slow and readily controllable. This important safety feature eliminates the need for an emergency core cooling system. Only a residual heat removal system is required for the long-term decay heat, and control of the HTGR is inherently easier than in reactors in which the coolant functions as the moderator. The uranium/thorium fuel contained in the ceramic-coated particles is not susceptible to sudden release of the stored-up fission products as a result of melting. Since the entire primary coolant system is contained within the PCRV, external piping, which might be subject to sudden rupture, is eliminated. Structural strength and integrity of the PCRV is enhanced by the redundant reinforcing steel and prestressed wire tendons. At the maximum credible pressure, the prestressing elements are not stressed above levels experienced during their initial tensioning. As a result, sudden loss of coolant due to prestress failure is not credible.

Second-Generation HTGRs

In addition to upgrading the HTGR at the Fort St. Vrain nuclear power station, efforts have been underway for several years to make both larger and smaller gas-cooled high-temperature reactors. Smaller, modular units could provide the flexibility needed by the public utilities as they plan their expansions for projected increases in electricity requirements. Inherent safety, already a feature of the HTGR, would be enhanced because of the smaller size and low power density of modular units. For example, it is estimated that the power density of a modular gas-cooled reactor would be only 3 kW/liter, as compared with 6 kW/liter for a large reactor and 100 kW/liter for a conventional pressurized-water reactor (PWR) as previously described.

In the new designs, if coolant were lost, the nuclear chain reaction would be terminated by the reactor's negative temperature coefficient after a modest temperature rise. Core diameter of the modular units would be limited so that decay heat could be conducted and radiated to the environment without overheating the fuel to the point where fission products might escape. Thus, inherent safety would be realized without operator or mechanical device intervention.

Large Commercial HTGRs[2]

Following construction of the Fort St. Vrain facility, the HTGR was marketed commercially in direct competition with large pressurized water reactors (PWRs) and boiling water reactors (BWRs). Between 1971 and 1975, ten such reactors were ordered by U.S. utilities. The designs of the commercial HTGR were similar to Fort St. Vrain in that they used the graphite based core structure, helium coolant, prestressed

[2]This portion of article on HTGRs contributed by R. A. Dean, Sr. Vice President, GA Technologies Inc., San Diego, California.

concrete reactor vessel (PCRV) and superheated steam cycle. However, the designs differed in that power outputs were significantly larger and the reactor system was rearranged to accommodate the larger-size components.

The large HTGRs had power ratings of 2000 and 3000 MWt which corresponded to net electrical outputs of 770 and 1160 MWe, respectively. An example of the rearranged reactor system is shown in Fig. 19. A multi-cavity PCRV was used to enclose the reactor system instead of the single-cavity PCRV used in Fort St. Vrain. This was a major advancement in PCRV technology and necessitated the development of a circumferential wire-wrap prestressing system instead of circumferential tendons, although the longitudinal tendons were retained. The sizes of the multi-cavity PCRV were approximately 100 feet (30 m) high by 120 feet (36 m) in diameter for the 3000 MWt plant and 100 feet (30 m) high by 105 feet (32 m) in diameter for the 2000 MWt plant.

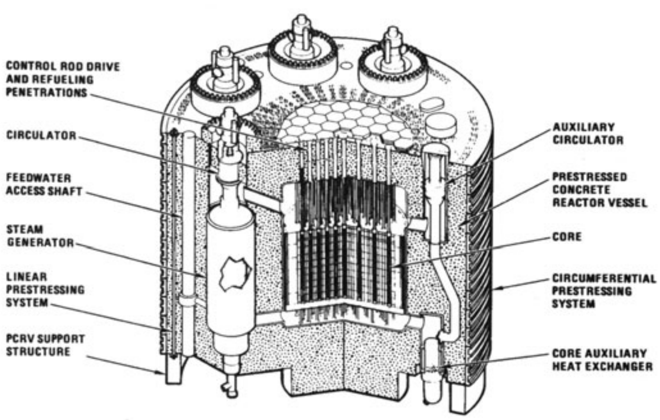


Fig. 19. Integrated HTGR nuclear steam system (1170 to 3360 MW thermal). (*GA Technologies.*)

The graphite reactor core was located in the central cavity of the PCRV. The steam generators and steam-driven main helium circulators were located in vertical cavities arranged around the periphery of the core. The 2000 MWt system had four steam generator-circulator side cavities while the 3000 MWt unit had six such cavities. The hot primary coolant helium (1366°F; 741°C) exiting from the bottom of the core collected in the lower core plenum from which it was distributed to the steam generators through the lower cross-ducts. The circulators, located above the steam generators, returned the cool helium (710°F; 377°C) through the upper cross-ducts to the upper core plenum. The helium in the upper plenum then flowed down through the core where it was heated.

Another feature of the large HTGRs was the core auxiliary cooling system which provided an independent means of core afterheat removal in the event that the main coolant loops (i.e., steam generators and steam-driven circulators) were shut down. The core auxiliary cooling system consisted of two redundant cooling loops, each capable of removing 100% of the afterheat, for the 2000 MWt plant and three redundant cooling loops, each capable of removing 50% of the decay heat, for the 3000 MWt plant. Each loop contained a motor-driven circulator and water-cooled heat exchanger and circulated flow through the core just at the main loops. Shutoff valves were located in both auxiliary and main loops to assure that helium would not bypass the core through the one system while the other system was in operation.

All ten large commercial HTGRs, ordered during the early 1970s as an indirect consequence of the energy crisis brought about by the oil embargo, were cancelled by 1976. The combination of the recession plus new emphasis on conservation brought about a rapid reduction in electric energy demand which, in turn, resulted in cancellation of over 100 nuclear power plant orders, including the large HTGRs.

Small Modular HTGRs

In the early 1980s, the major influence on new designs was the renewed emphasis on safety brought about by the accident at the Three Mile Island nuclear plant. The experience from licensing and operation of nuclear plants during the 1970s indicated a need to reduce design complexity and develop passive approaches to reactor safety rather than rely on complex emergency safety systems. HTGR designers in the United States and Europe determined that a substantial reduction in plant size could enable the HTGR to be entirely inherently safe by virtue of the high temperature structural integrity of the graphite core and ceramic coated fuel. This means that a small HTGR would not require any active safety equipment or any action by the operator in order to prevent release of radioactivity for any accident condition.

A major concern with reducing plant size was the economic impact from reversing the economy of scale. However, it was learned that economy-of-scale effects could be offset by several beneficial factors which apply to smaller nuclear plants. This includes the shift of major portions of the work from the site to the factory; the learning effects appreciated by replication of a larger number of smaller units in a factory environment and the elimination or simplification of many components/systems no longer required for smaller plants.

These considerations led to the reconfiguration of the HTGR plant into a system of one or more downsized 350 MWt modular reactors. The physical arrangement of a single reactor module, designed for installation in a below-grade silo, is shown in Fig. 20. The primary components are contained within two vertically oriented metal pressure vessels connected by coaxial cross-duct. Thus, the field-erected PCRV, which was used on previous large HTGRs, was eliminated in favor of shop-fabricated metal pressure vessels. The use of metallic pressure vessels also facilitates installation in underground silos, which enhances the safety of the plant.

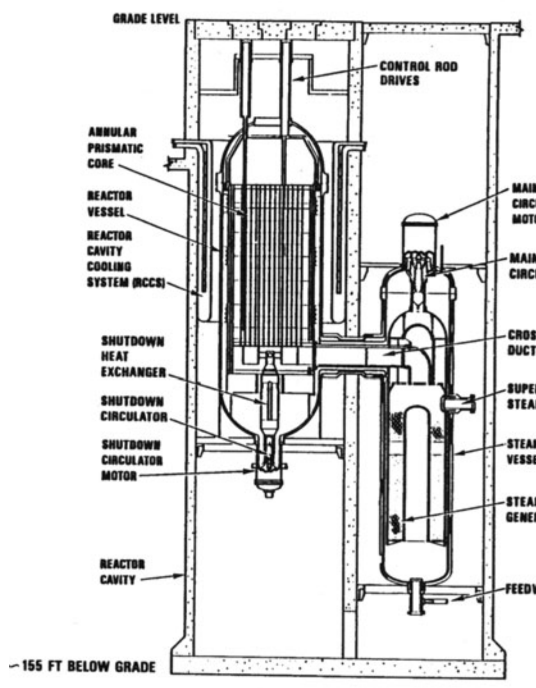


Fig. 20. Elevation of 350 MW modular HTGR. (*GA Technologies.*)

The reactor vessel, which is approximately 72 feet (22 m) high by 22.5 feet (6.9 m) in diameter, contains the graphite core, reflector, and shutdown heat removal system (non-safety). The other vessel contains a single helical coiled steam generator and motor-driven circulator with magnetic bearing instead of water-lubricated bearings as in previous concepts. The size of both vessels is within allowable limits for barge, rail, and overland transportation.

During normal operation, the main circulator transports hot helium at 1266°F (686°C) from the bottom of the core to the steam generator which, in turn, produces superheated steam at 1005°F (541°C) and 2500 psia. The cold helium at 496°F (258°C) is returned to the top of the reactor core. During normal shutdown and refueling, the non-safety auxiliary shutdown heat removal system removes core afterheat if the main heat transport system is not operational.

A principal feature of the modular HTGR is its capacity for safely rejecting core afterheat in a completely passive manner (i.e., without the need for any active core cooling systems) such that any release of fission products from the fuel is prevented during severe accident conditions. This feature is a result of both the reactor system configuration and the high temperature capability of the fuel. In the event of a loss of forced circulation cooling of the core via either the main circulator/steam generator or auxiliary shutdown circulator/heat exchanger, core afterheat will continue to be safely removed by direct conduction through the core and reflector to the reactor vessel wall. The heat is then dissipated from the reactor vessel surface by radiation and natural convection to cooling panels surrounding the interior surface of the reactor cavity. See Fig. 20, previously mentioned. These panels are part of the Reactor Cavity Cooling System (RCCS) which consists of natural convection air ducts that ultimately transport the core afterheat directly to the atmosphere.

In order to achieve this passive core cooling capability in a reactor with a power level and power density that are economically attractive, the annular core arrangement shown in Fig. 21 was adopted. The active core consists of an annular region of hexagonal graphite fuel blocks containing standard HTGR fuel. Unfueled graphite reflector blocks make up the region inside the active core annulus and the region surrounding the outside of the annulus. This arrangement results in a higher radial heat conductance for fuel at the innermost radius than for a solid cylindrical active core. Thus, the annular core arrangement permits operation at a higher power for a given volume than a solid active core.

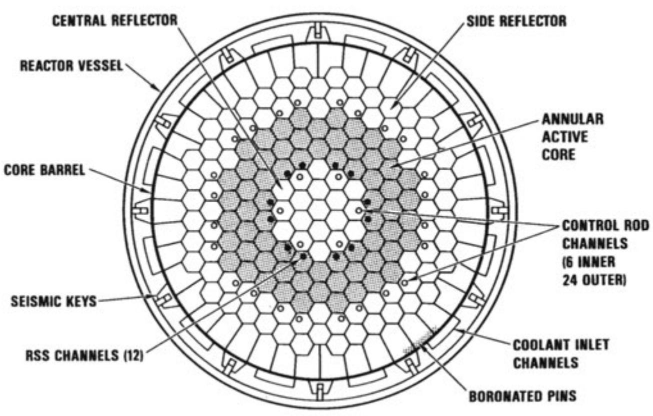


Fig. 21. Reactor core cross section of 350 MW modular HTGR. (*GA Technologies.*)

The modular HTGR uses the same form of fuel as the large HTGR and Fort St. Vrain installation except for the important difference that the ^{235}U enrichment was reduced to about 20% from the previous value of 93%. The fuel is in the form of coated fuel and fertile particles which are bonded into graphite rods and inserted into the hexagonal graphite fuel blocks. See Fig. 22. The fuel and fertile particles consist of uranium oxycarbide and thorium oxide kernels (about 350 micrometers in diameter), respectively, first coated with a porous graphite buffer, followed by three successive layers of pyrolytic carbon, silicon carbide, and pyrolytic carbon. The outer diameter of the coated particles is about 800 micrometers for the uranium particles and slightly larger for the thorium particles. The coatings essentially form a high-temperature refractory-based pressure vessel around each fuel/fertile kernel for the purpose of retaining fission products. Extensive operation and test data on these particles confirm that essentially no failure of the refractory coating occurs if the fuel is maintained below 3272°F (1800°C). As previously mentioned, the reactor design parameters were selected such that passive core afterheat removal will prevent this temperature from

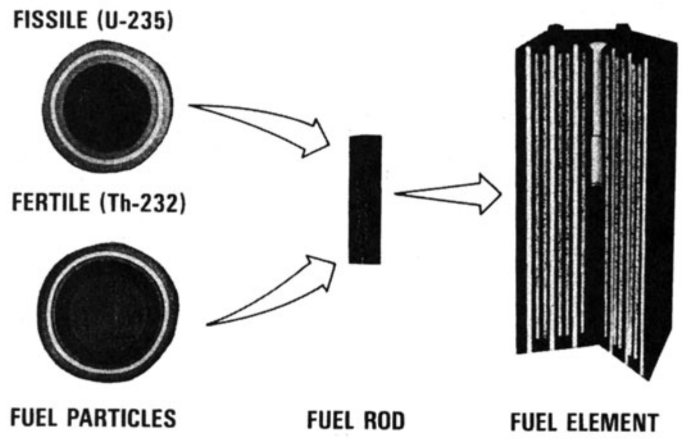


Fig. 22. Fuel components of HTGR. (*GA Technologies.*)

The reference plant arrangement features four 350 MWt HTGR modules supplying steam to two turbine generators that produce a net electrical output of 558 MWe at a new plant efficiency of 39.9%. Each reactor module is housed in an independent, vertical, cylindrical, concrete confinement, which is fully embedded in the earth. The four reactor modules share common systems for fuel handling, helium processing, and other essential services. A common control room is used to operate all four reactors and the turbine plant. Operation of the entire complex is completely automated. Human operator actions are not required for control during power production or to assure safe shutdown during hazardous conditions.

Potential for HTGRs

The HTGR's use of ceramic-coated fuel and graphite moderator enables operation of the reactor core at much higher temperatures than are required for electric power production via a steam Rankine cycle. Core outlet helium temperatures in excess of 1800°F (982°C) are achievable without impacting the integrity of the HTGR fuel or core structures. This very high temperature capability opens up the possibility for more efficient methods of power production or direct use of high-temperature thermal energy for process heat applications. See also **Cogeneration.**

An attractive electric power producing concept for the 21st century is the HTGR gas-turbine (HTGR-GT) which has the potential for thermal efficiencies over 50% by taking advantage of the high HTGR core outlet temperature. Although several variations in system configuration are possible, the most straightforward HTGR-GT concept is the direct Brayton cycle, illustrated in Fig. 23. This cycle is closed and the helium primary coolant is also the working fluid for the power conversion system. The entire heat source and power conversion system of an HTGR-GT, which is capable of a net electric output of 170 MWe, can be enclosed within the two pressure vessels and cross-duct arrangement similar to the modular steam cycle HTGR previously shown in Fig. 20.

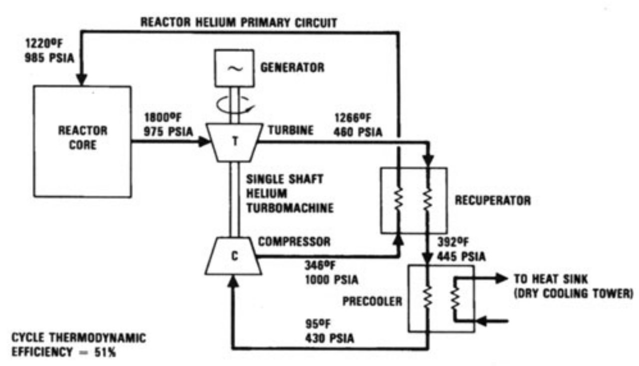


Fig. 23. HTGR gas turbine system with exceptional cycle thermodynamic efficiency. (*GA Technologies.*)

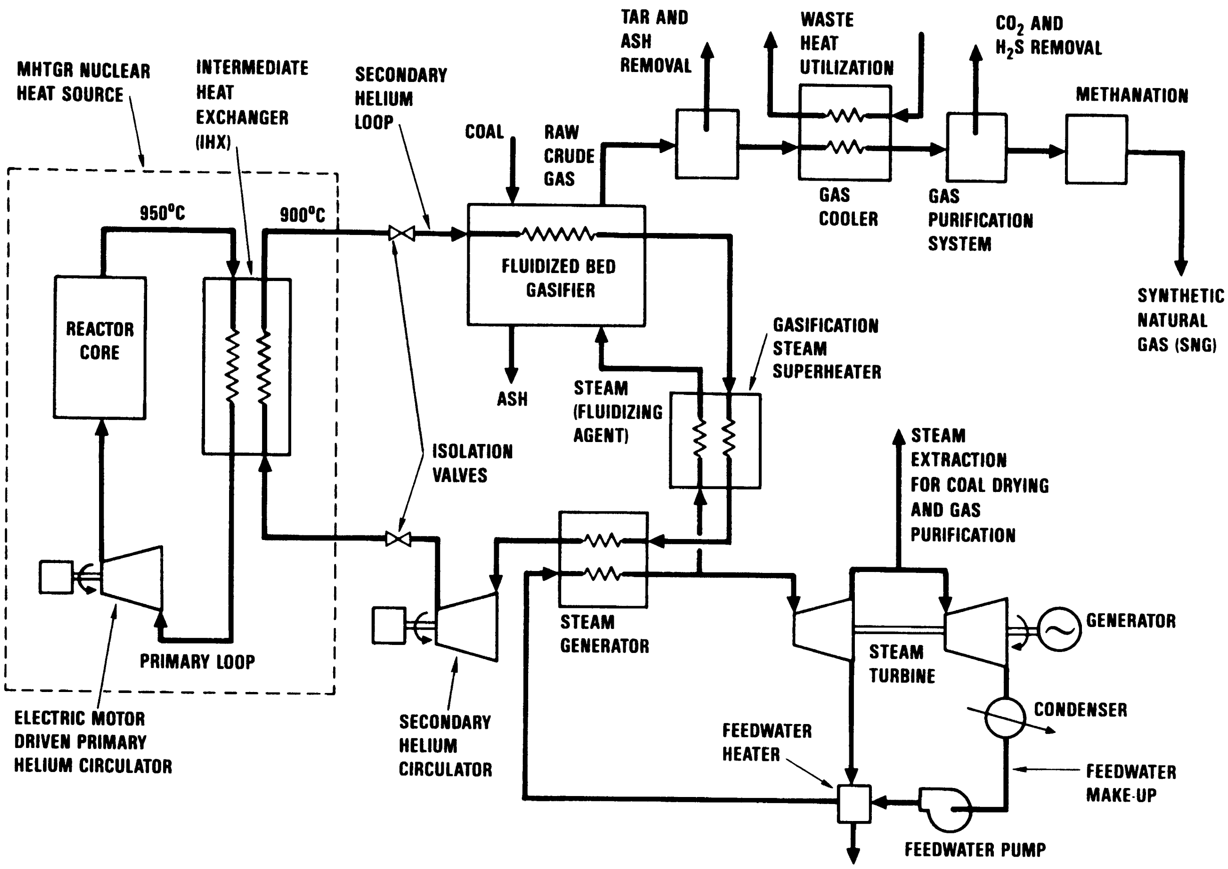


Fig. 24. Advanced process heat HTGR with intermediate loop for producing synthetic natural gas (SNG) by steam gasification of coal. (*GA Technologies.*)

The recuperator and precooler would occupy the same space as the steam generator and the turbo-compressor would replace the circulator.

Perhaps the most significant potential use of the HTGR's high temperature capability is the production of synthetic fuels from coal. The HTGR is an important option for supplanting the current consumption of oil and natural gas with synthetic natural gas (SNG). The HTGR can supply the necessary energy for this endothermic process and, therefore, increase the recoverable energy in the SNG product by at least 60% over traditional coal combustion processes.

The most effective method of SNG production with an HTGR is the steam-carbon reforming process in which superheated steam reacts with pulverized coal to form methane-rich SNG. A system for accomplishing this process is shown in Fig. 24. In this system, an intermediate heat exchanger (IHX) has been used to isolate the nuclear heat source from the process steam, thus allowing the use of conventional equipment for the SNG production portion of the plant. The IHX and reactor can be configured in the same arrangement as previously shown in Fig. 23, except that the IHX would occupy the space allocated to the steam generator.

Both gas-turbine and process heat versions of the HTGR are based on the demonstrated high-temperature capability of the fuel and core structure. However, some development in the metallic components, such as the turbine, hot ducts and intermediate heat exchanger is necessary. Present commercial alloys would have limited lifetime under service conditions at 1650°F (899°C) and above. However, currently envisioned advancements in ceramics and carbon-carbon composites indicate that high-temperature nonmetallic substitutes for metallic alloys will soon be available. These materials advances are the key to making future application of the HTGR a reality.

HEAVY WATER REACTOR (HWR)

During the atomic energy developments in the World War II years and for a period thereafter, the United States, the United Kingdom, and Canada cooperated closely and many of the nuclear scientists of these countries appreciated the merits of heavy water as a moderator. Each of these countries pursued some development of HWRs for commercial power generation, but at different paces and dedication. Only Canada took to the HWR for commercial power generation. See Figs. 25 and 26.

One of the first high-priority nuclear applications of the United States was for naval propulsion. Because of a very tight minimal physical size criterion, LWRs offered advantages over the HWR. The United Kingdom placed emphasis on the production of plutonium for weapons programs. Gas graphite reactors were a reasonable early choice. When commercial nuclear power was recognized as a needed source of energy, because of the accumulated operating experience it was reasonable to adapt the reactors which had already been developed in the United States and the United Kingdom for military purposes. Long-term savings at that time was not a major criterion.

In the postwar years, hydroelectric power amply met a large portion of Canada's power needs and its abundance made nuclear power quite noncompetitive. Canadian utility operators were used to capital-intensive plants combined with low operating costs. In analyzing the prospects for nuclear power in Canada, utility planners and engineers placed a significant value on low fueling costs, and thus neutron economy was paramount. Therefore, when commercial nuclear power studies commenced in Canada in the mid-1950s, the choice was the HWR. This choice was bolstered by experience with and knowledge about heavy water production plants gained when Canadian scientists were trading experience from the heavy water-moderated NRX research reactor when the United States was developing the Savannah River production facility, which was dismantled in the early 1990s.

Principal advantages of heavy-water reactors are: (1) more efficient absorption of the energy released in the reactor, (2) greater fuel "burn-up" and, therefore, fuel economy, and (3) refueling can take place while the reactor is in service.

The first Canadian nuclear power demonstration (NDP) reactor was of 20-MWe capacity and was configured similarly to a light water reac-

Fig. 25. Series of towers comprising part of the heavy water production plant at Ontario Hydro's Bruce nuclear power complex near Tiverton on the shores of Lake Huron. Heavy water is a clear, colorless liquid that looks and tastes like ordinary water. It occurs naturally in ordinary water in the proportion of approximately one part heavy water to 7000 parts of ordinary water. While ordinary water is a combination of hydrogen and oxygen (H_2O), heavy water (D_2O) is made of up of deuterium—a form, or isotope, of hydrogen—and oxygen. Deuterium is heavier than hydrogen in that it has an extra neutron in its atomic nucleus, so heavy water weighs about 10% more than ordinary water. It also has different freezing and boiling points. It is the extra neutron that makes heavy water more suitable than ordinary water for use in CANDU nuclear reactors as both a moderator and a heat transport medium. (*Ontario Hydro, Toronto, Ontario, Canada.*)

tor. Because of limited facilities for making large pressure vessels, a modular pressure-tube design of the configuration shown in Fig. 24 was investigated. Zircaloy-2 had become available at that time for fabrication of the pressure tubes. Hence the NPD was constructed using Zircaloy as cladding material and uranium dioxide as fuel. The NPD reactor has been in operation since 1962. The CANDU (Canada Deuterium Uranium) power reactors, including the NPD, number over twelve facilities. See Figs. 27, 28, and 29.

CANDU power reactors are characterized by the combination of heavy water as moderator and pressure tubes to contain the fuel and coolant. Their excellent neutron economy provides the simplicity and low costs of once-through natural uranium cycling. Future benefits include the prospect of a near-breeder thorium fuel cycle to provide security of fuel supply without the need to develop a new reactor, such as the fast breeder. The CANDU system is appropriate for countries of intermediate economic and industrial capacity, such as Canada. Producing heavy water is fundamentally simpler than enriching uranium and commercial heavy water plants have been built in smaller sizes than would be possible for uranium enrichment plants. Although Canada has rather generous supplies and reserves of uranium, there is increasing pressure on Canada to export uranium, a pressure that will probably intensify further if the introduction of fast breeder reactors in other countries is delayed. The current simplest possible fuel cycle for the CANDUs, which is not dependent upon fuel reprocessing, will probably be retained in Canada so long as uranium remains plentiful and comparatively economical. However, for future planning, research to date has indicated that a "self-sufficient thorium cycle" may be practicable in the CANDUs with minimal modification. It has been observed that, at equilibrium, the thorium cycle would require no further uranium. Only small quantities of thorium, which is more abundant than uranium, would be required. Also of interest for the future is *electronuclear breeding*, i.e., the use of electric power to convert fertile to fissile material for neutron economy.

FAST BREEDER REACTORS

The fast breeder reactor derives its name from its ability to breed, that is, to create more fissionable material than it consumes. This ability stems from the fact that neutrons travel faster than they do in a thermal reactor. The breeding process depends, in part, upon the neutrons maintaining a high speed, or high energy. If their speed or energy is allowed

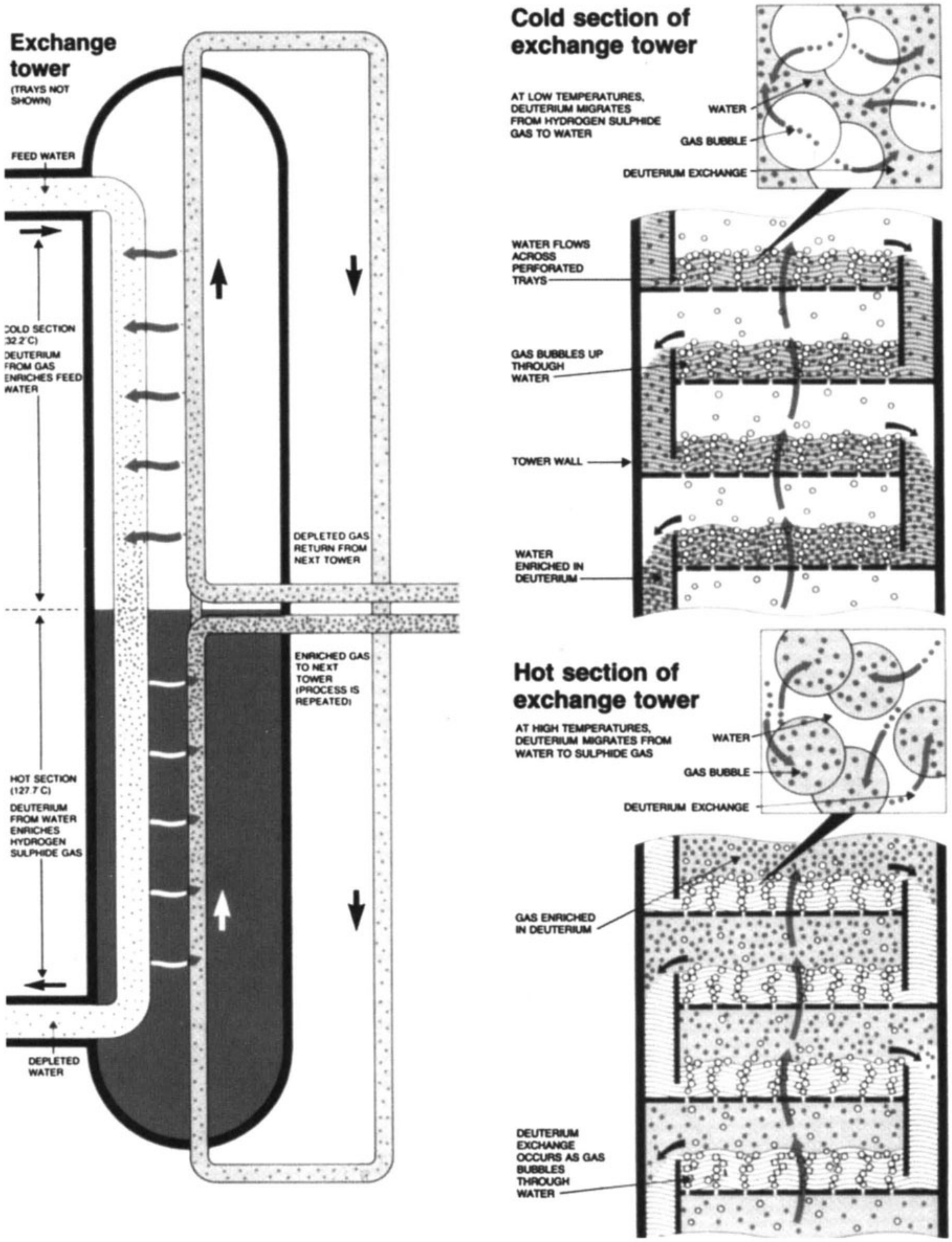


Fig. 26. The production of heavy water is based upon the behavior of deuterium in a mixture of water and hydrogen sulfide. When liquid H_2O and gaseous H_2S are thoroughly mixed, the deuterium atoms exchange freely between the gas and the liquid. At high temperatures, the deuterium atoms tend to migrate toward the gas, while they concentrate in the liquid at lower temperatures. In the first and second stages of production, the towers of a heavy water plant are operated with the top section cold and the lower section hot. Hydrogen sulfide gas is circulated from bottom to top and water is circulated from top to bottom through the tower. In the cold section, the deuterium atoms move toward the water and are carried downward, while in the hot section, they move toward the gas and are carried upward. The result is that both gas and liquid are enriched in deuterium at the middle of the tower. A series of perforated trays are used to promote mixing between the gas and water in the towers. A portion of the H_2S gas, enriched in deuterium, is removed from the tower at the juncture of the hot and cold sections and is fed to a similar tower for the second stage of enrichment.

The first stage of the process enriches the gas from 0.015% deuterium to 0.07%. A second stage further enriches it to about 0.35%. Again, the enriched gas is fed forward to a third stage. The product from this third stage, now in the range of 10 to 30% heavy water, is sent to a distillation unit for finishing to 99.75% purity "reactor-grade" heavy water. Because the production of heavy water uses a toxic gas, H_2S, safety is a top priority at heavy water plants. H_2S is a colorless gas, slightly heavier than air. To safely expel H_2S from the system, it is directed to a flare tower where it is burned off. Initially, each of Ontario Hydro's reactors requires about 800 megagrams, or one year's production from a heavy water plant. After that, less than 1% of the heavy water is lost and has to be replenished each year. (*Ontario Hydro, Toronto, Ontario, Canada.*)

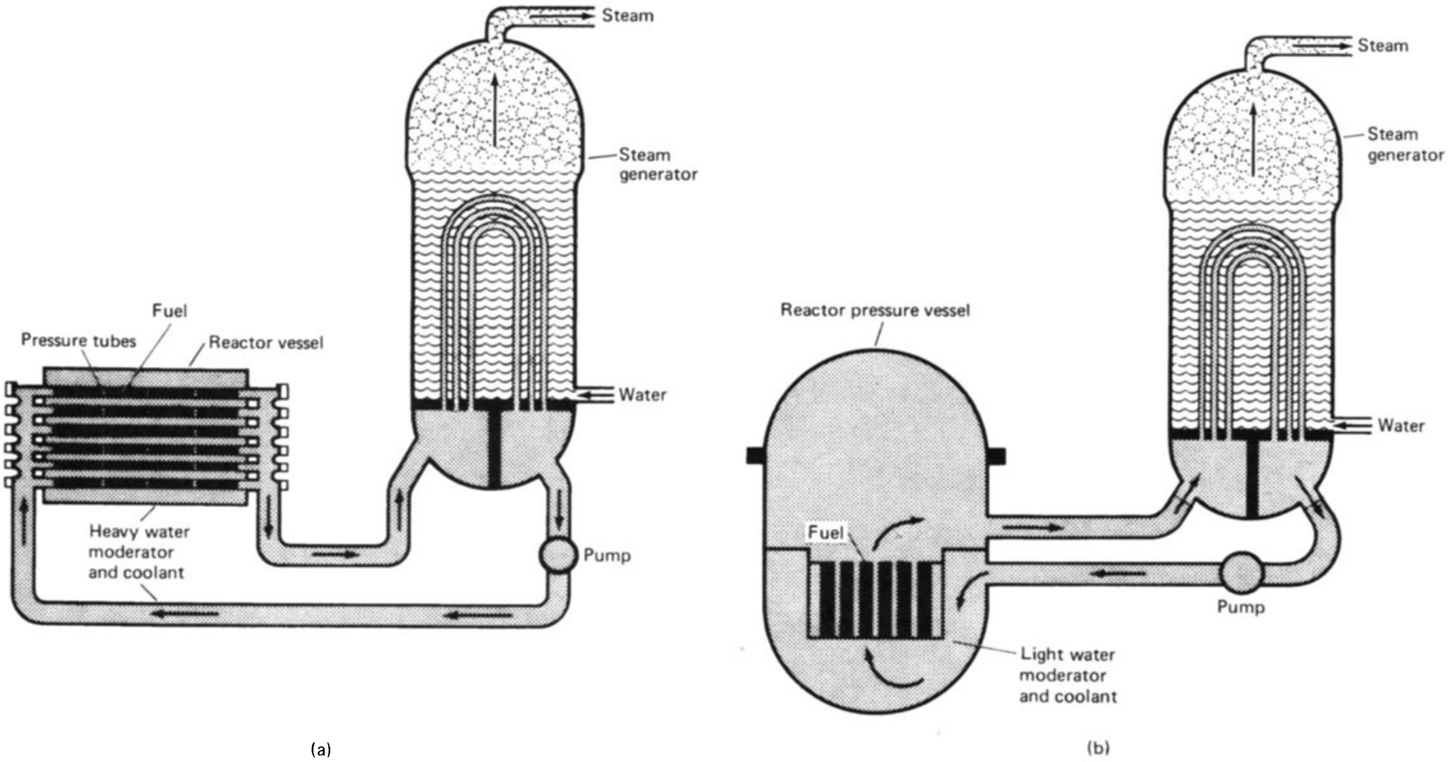


Fig. 27. Comparison of heavy water reactor (a) with light water reactor (b). (*After Robertson, Atomic Energy of Canada Limited.*)

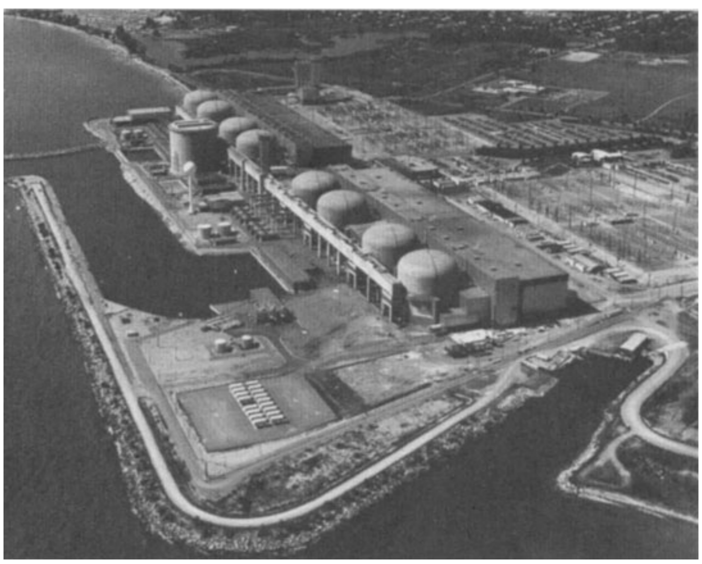


Fig. 28. Pickering (Ontario) "A" generating station's initial performance was outstanding and the station was hailed as a major Canadian technological achievement at the time of its commissioning in February 1971. It reached full power in 3 months, well ahead of schedule. The final Unit 4 went on line in May 1973. At full output (2,160,000 kW), Pickering "A" generates enough power to supply more than 1.5 million homes. In 1974, construction was begun on a twin station, Pickering "B," also shown in this view. The "B" station has the same capacity as its forerunner and all four units became operational in 1986. (*Ontario Hydro, Toronto, Ontario, Canada.*)

to degrade as occurs in thermal reactors, the number of neutrons produced per absorption in uranium or plutonium decreases. Furthermore, at lower velocities, neutrons tend to be captured in various structural materials of the reactor, and this further reduces the breeding potential. It is important, therefore, in fast reactors to keep the velocity of the neutrons high. Water, which is used as a coolant in some thermal reactors, tends to slow the neutrons down and thus prevent efficient breeding. Therefore, it is necessary to use a coolant which does not slow the

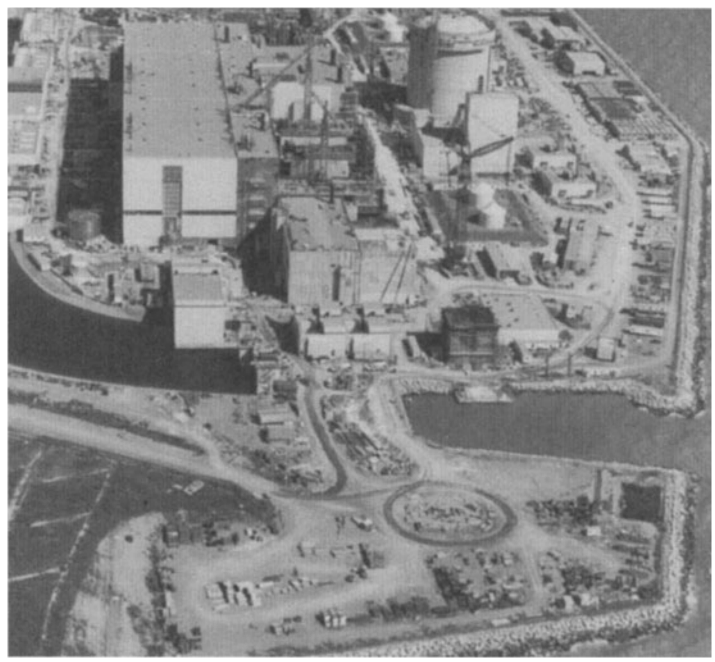


Fig. 29. Darlington, Ontario, nuclear generating station, one of the largest energy projects undertaken in North America. The plant is shown in late stages of construction in 1986. It is located on the shores of Lake Ontario, 5 km southwest of Bowmanville in the Town of Newcastle. Currently nuclear power plants provide one-third of Ontario's electricity needs (the other two-thirds comes from hydro installations). Selected because of its proximity to the residential, industrial, and commercial energy markets of Ontario, the 1200-acre (485-hectare) site has good transportation access, an abundant supply of cooling water from Lake Ontario, relative isolation, and excellent bedrock for station foundations. When the four-unit station is completed it will provide 3,524,000 KW of electricity, enough to serve a city of 3 million people. Electricity from the four units (each 881,000 KW net) will be fed into the Ontario Hydro electricity grid system through 500,000 V transmission lines already crossing the site. The Darlington generating station was approved by the provincial government in July 1977. Site preparation began in 1978 and by 1981, the first concrete was poured for the station's foundations. Construction activity peaked in 1986–1987 with approximately 6800 workers on site. (*Ontario Hydro, Toronto, Ontario, Canada.*)

neutrons or capture them as they travel through the coolant. Liquid sodium and gaseous helium under pressure are the two principal coolants used to date.

Fuel Cycle Considerations. Approximately 99.3% of uranium as it is found in nature is the isotope ^{238}U and 0.7% is ^{235}U. Uranium-235 is a fissile isotope, that is, if it is struck by a neutron it will split, or fission yielding on the average approximately two neutrons and 200 MeV of energy. This amount of energy corresponds to approximately 78 million Btu for every gram of uranium which fissions (3.5×10^{10} Btu/pound) (1.95 Calories/kilogram). Most reactors which exist today are largely dependent upon ^{235}U for their energy. However, some of the neutrons released in fission of ^{235}U also are absorbed in nonfissionable ^{238}U. As the ^{238}U absorbs a neutron, it is transformed into fissionable ^{239}Pu (plutonium). Thus, while the reactor is sustaining the fission process and thereby creating energy, it is also generating fresh fuel which can later be used to create more energy. Unfortunately, this is an inefficient process in present thermal reactors where the neutron velocity is established by the temperature, or thermal energy, so only limited amounts of additional energy are made available by transformation of ^{238}U into ^{239}Pu.

The fast breeder reactor makes possible the recovery of most of the available energy in uranium. This occurs because during fission in the fast breeder nearly three neutrons are released for every neutron absorbed as compared with only approximately two neutrons in a thermal reactor. On the average, between one and two neutrons are necessary for sustaining the fission process, and the extra neutron in a fast reactor can be absorbed in nonfissionable ^{238}U and thereby transformed into fissionable ^{239}Pu. Reactors which have a breeding ratio greater than one create more fuel than they need for their own purposes, and the extra plutonium can be used to fuel new breeder reactors. By this means, 80% or more of the available energy in uranium can be recovered and used in reactors.

In a typical fast breeder, most of the fuel is ^{238}U (90 to 93%). The remainder of the fuel is in the form of fissile isotopes which sustain the fission process. The majority of these fissile isotopes are in the form of ^{239}Pu and ^{241}Pu, although a small portion of ^{235}U can also be present. Normally, the fissile isotopes are located in a central "core" region which is surrounded by the fertile isotopes in the "blanket" region. This is illustrated in Fig. 30.

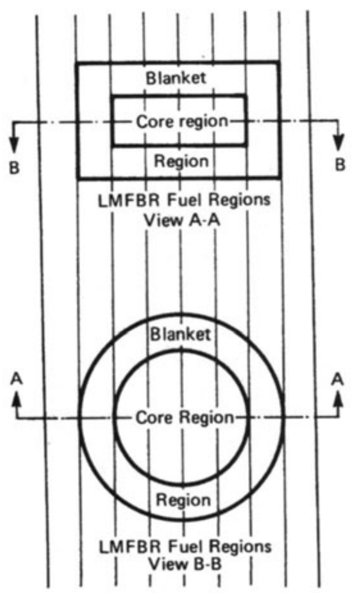


Fig. 30. Liquid-metal fast breeder reactor core and blanket arrangement.

When the fuel is initially loaded into the reactor, the core region will typically contain from 10 to 15% fissile isotopes with the remainder being ^{238}U. Essentially all of the blanket will be ^{238}U. As energy is extracted from the fissile isotopes, they become depleted (the initial plutonium is gradually used up). However, in a breeder reactor, new plutonium will be formed in the core and blanket regions faster than it is consumed. Additionally, undesirable fission products are formed which must ultimately be removed. This process is schematically illustrated in Fig. 31. The "before" chart represents the new fuel condition and the "after" chart corresponds to the situation when the fuel is removed for reprocessing. Typically, the fuel which is removed for reprocessing will contain from 1 to 3% new plutonium. It is in this manner that the fast breeder can recover from 80 to 90% of the available energy in uranium resources. Most present reactors require some enrichment of the ^{235}U isotope used to fuel them. This enrichment process requires a plant which, in turn, uses large amounts of electrical energy. Because the fast breeder converts the fertile isotope ^{238}U into the fissile isotope ^{239}Pu, no enrichment plant is necessary. The fast breeder serves as its own enrichment plant. The need for electricity for supplemental uses in the fuel cycle process is thus reduced.

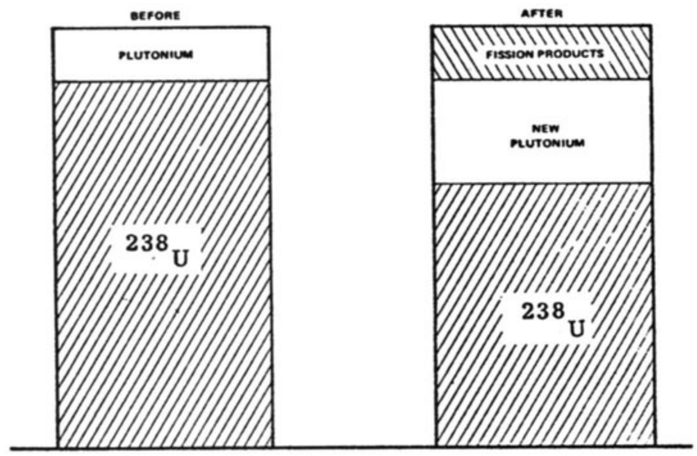


Fig. 31. Basic operation of the breeder reactor. The illustration does not include geometrical disposition of fuel in the core and blanket system. (*General Electric*.)

Fast Breeder Reactors in Perspective

Of the several fundamental ways to use nuclear fission reactions to generate electric power, the fast breeder reactor (notably the liquid-metal-cooled fast breeder reactor, LMFBR) probably has the most checkered history. The fast breeder reactor received its early impetus when there was serious concern over what appeared to be a limited supply of uranium and consequent increasing prices of uranium fuels. With the fast breeder concept offering up to a 100-fold increase in the utilization efficiency of uranium, it appeared to be the logical replacement (second generation) for light-water reactors. It would present a technical solution in time for the expected tight supplies of uranium. The United States funded LMFBR research quite generously until a serious reduction in 1982, when it was determined that the shortage of uranium fuels no longer posed a serious threat in the short-term, thus establishing the general consensus in the U.S. that the breeder, if needed at all, could be delayed until well into the next century. The interest of the French and Japanese in breeders also stemmed from early concerns with uranium shortages and prices, but was much more serious because these countries are extremely uranium poor and, further, these countries have fewer energy options. For example, the cost of coal ranges from 1.6 to 2.5 times the cost of coal in the United States. Continued progress in fast breeder development, particularly in France, also has been accelerated by a very heavy past investment in the technology coupled with a desire to be fully self-contained as regards the generation of electric power.

Particularly, in the United States because of its several energy fuel options and its current "rethinking of nuclear power," the principal nuclear power research targets no longer include uranium supplies, but rather the lowering of capital costs for existing light-water reactor technology, shortening the plant construction and licensing lead times, achieving higher plant availability (essentially eliminating long power outages), and increasing plant safety. From this "rethinking" process, the U.S. Congress suspended funding for the Clinch River Breeder Reactor.

It is interesting to note that, as of the early 1990s, most authorities are not quite willing to "forget" the fast breeder altogether. The differences in views essentially reside in timing. The fast breeder, even though costly to build today, does offer several temptations. For example, experience gained from the Experimental Breeder Reactor (Idaho Falls, ID) which commenced operation in 1964 is reported to operate better now than when it was first built, showing no evidence of corrosion. It is suggested that if such experience were extrapolated the useful life of an LMFBR could be between 100 and 150 years! Weinberg (1986 reference listed) recalls, "that one of Newcomen's original steam engines, built in the mid-18th century, continued to operate until 1918." It does appear that the economy of the LMFBR is somewhat analagous to large hydroelectric projects that require very large investments of capital, but coupled with low operating costs, provide really cheap electric power once they are fully amortized. However, one usually must wait a generation before this situation occurs. Weinberg also observes, "Because the breeder requires little, if any, mining of uranium, its environmental impact is much smaller, at least at the front end of the fuel cycle, than is the impact of the LWR. The roughly 300,000 tons of depleted uranium stored outside the diffusion plants, if used in breeders, could fuel our (U.S.) entire electric system for centuries!"

Davis (1984 reference listed) observes that there is sufficient know-how today to build and operate fast breeder reactors with confidence of their safety and reliability, as exemplified with large units in France. Davis further suggests that the principal problems remaining in fast breeder technology include: (1) a reduction in overall costs to make the fast breeder competitive with coal and the light-water reactors—an estimated reduction factor of 1.5 to 2; (2) assurance of adequate safeguard on the plutonium fuel cycle; (3) more demonstrations of commercial-scale operations, such as the engineering scale-up of important system components—pumps and steam generators; (4) implementing a large, overall system which must include parallel facilities for fuel processing and refabrication; and (5) setting in place the reprocessing of light-water reactor fuel to provide the "start up" plutonium for the fast breeders.

Some technical successes with the fast breeder have occurred in France: (1) demonstrations of a positive breeding gain of 0.15 ± 0.04 in a complete breeder fuel cycle, and (2) demonstration at several laboratories of uranium and plutonium oxide fuel elements that can sustain more than 10% burnup of the original mixture of $^{238}U/^{239}Pu$ before the fuel has to be reconstituted, representing a tenfold improvement (Weinberg, 1986). It has also been observed that, compared with a number of so-called alternative fuels, the new Super-Phoenix (France) plant can produce electricity at costs that are markedly less than electrical energy from solar-powered photovoltaics, for which funding remains significant. Another proposed inexhaustible power source, nuclear fusion, still remains in an early stage of development. See entry on **Fusion Power.**

LMFBR Design Principles. There are many design differences among the reactor designs, including: (1) primary coolant system arrangement; (2) refueling mechanism design and arrangement; (3) steam generator type and arrangement; (4) core support method; (5) structural material choices; and (6) safety features. Perhaps the most noticeable difference is that of the primary system arrangement. This difference is schematically illustrated in Figs. 32 and 33. The system of Fig. 32 corresponds to a "loop" or "piped" arrangement where the reactor, pumps, and intermediate heat exchangers are located separate from each other and piping carries the sodium from one point to the other. The "pool" or "tank" arrangement of Fig. 33 includes the reactor, intermediate heat exchangers and pumps in one large pool of sodium which is contained in a separate tank. Each concept has advantages and disadvantages. The pool concept is somewhat easier to design for certain hypothetical accident situations. The loop concept is easier to construct and to maintain.

The flow circuit for an LMFBR where two sodium circuits are included is shown schematically in Fig. 34. The reactor is cooled by the primary sodium, which becomes radioactive as it picks up heat in passing through the core or fueled region. In this particular arrangement, the sodium is heated to 560°C and flows through pipes (schematically shown as a single line in the figure) to the intermediate heat exchangers. In the heat exchangers, the primary sodium transfers heat to the nonradioactive sodium. After being cooled to 393°C in the heat exchangers, the primary sodium is pumped back into the reactor where it again repeats the circuit. The nonradioactive secondary sodium is circulated from the intermediate heat exchangers through

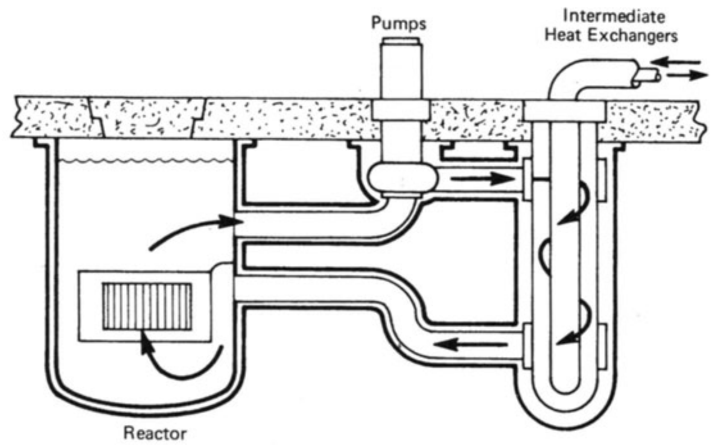


Fig. 32. Loop arrangement in the liquid-metal fast breeder reactor. (*General Electric.*)

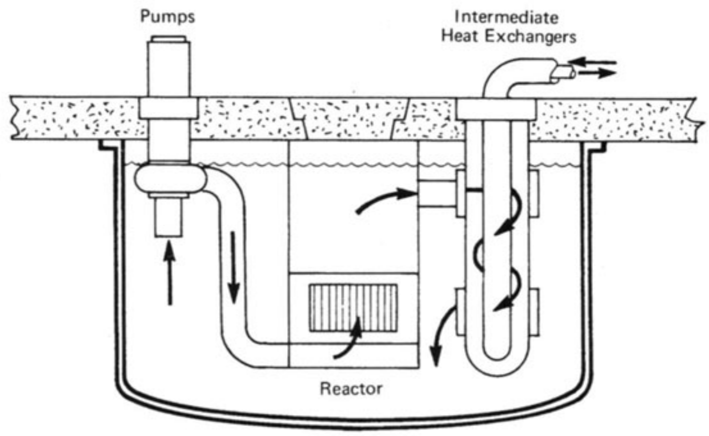


Fig. 33. Pool arrangement in the liquid-metal fast breeder reactor. (*General Electric.*)

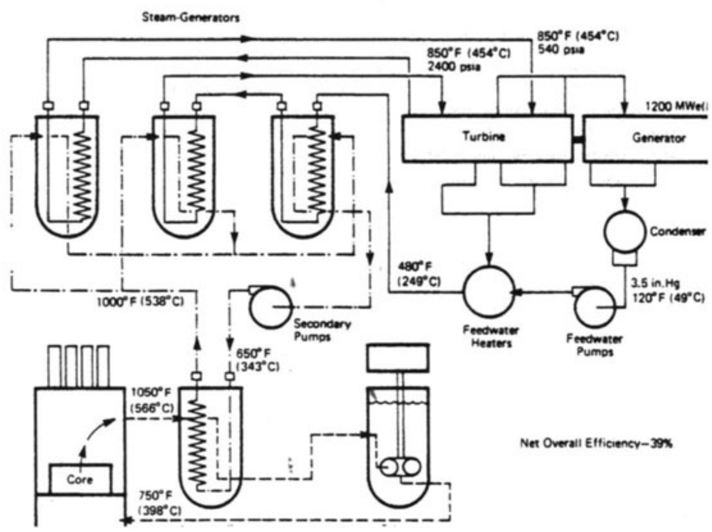


Fig. 34. Liquid-metal fast breeder reactor flow circuit. (*General Electric.*)

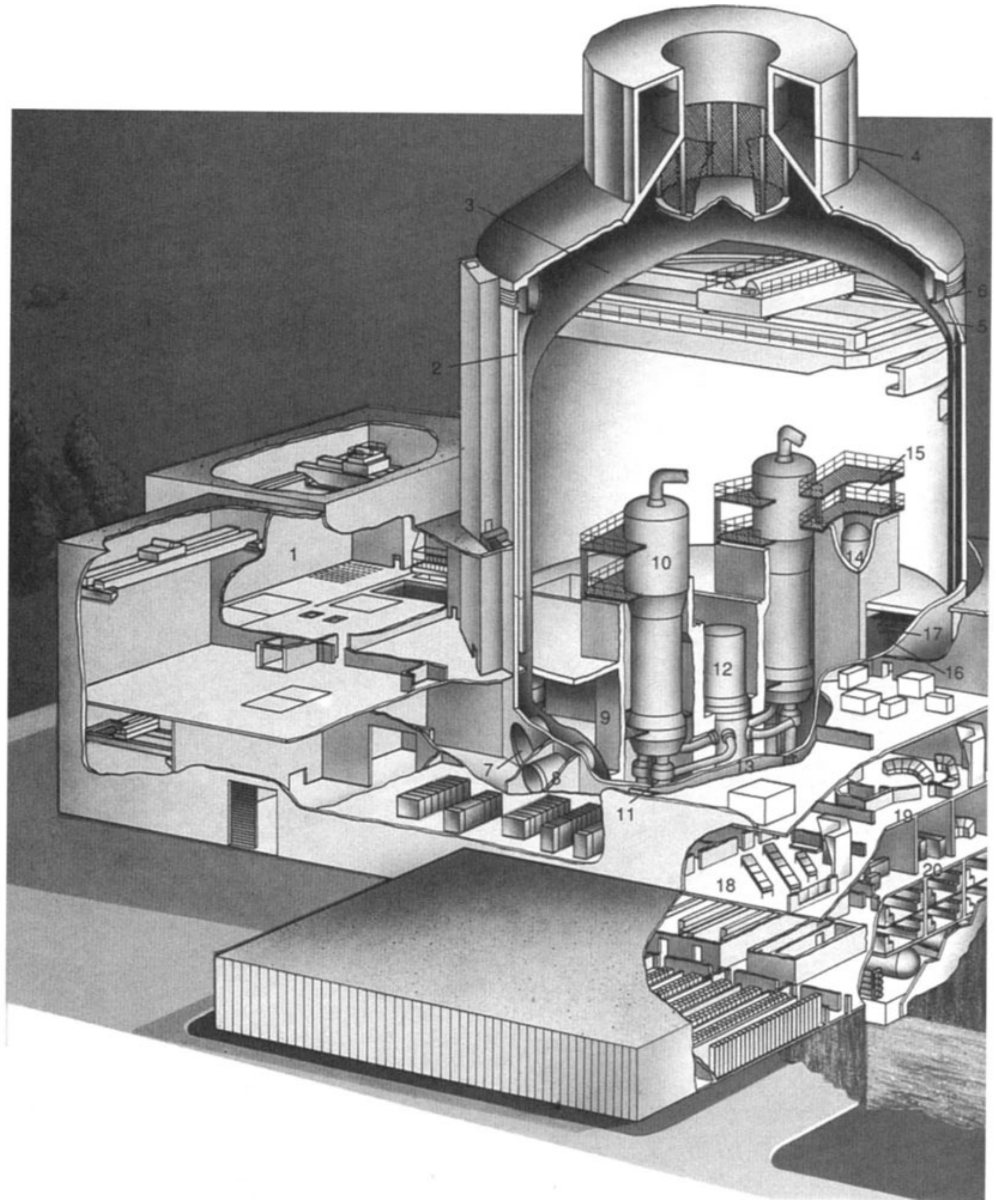


Fig. 35. Sectional view of the 600 MWe pressurized water reactor (PWR) expected to be operational by 1995. Considered the PWR of the future, the plant is designed for a minimum useful life span of 60 years and features numerous economic and safety features, including passive systems for ultimate protection. (*Joint project of Westinghouse, the Electric Power Research Institute, and the U.S. Department of Energy.*)

Legend:

1. Fuel Handling Area
2. Concrete Shield Building
3. Steel Containment
4. Passive Containment Cooling Water Tank
5. Passive Containment Cooling Air Baffles
6. Passive Containment Cooling Air Inlets
7. Equipment Hatches (2)
8. Personnel Hatches (2)
9. Core Make-up Tanks (2)
10. Steam Generators (2)
11. Reactor Coolant Pumps (4)
12. Integrated Head Package
13. Reactor Vessel
14. Pressurizer
15. Depressurization Valve Module Location
16. Passive Residual Heat Removal Heat Exchangers
17. Refueling Water Storage Tank
18. Technical Support Center
19. Main Control Room
20. Integrated Protection Cabinets

steam generators where the heat from the sodium is transferred to water, which becomes superheated steam for use in the turbine. The cooled secondary sodium is pumped back through the intermediate heat exchangers where the process is repeated. Steam from the steam generators is used to turn the rotor of the turbine generator to generate electricity. In the arrangement shown, 1,200 MW of electricity are generated at a net overall efficiency of 39%. This relatively high efficiency is possible because of the excellent thermal characteristics of sodium.

during the mid-1980s. Advanced programs were established to create new plant designs that would reflect past experience to achieve an extremely high degree of operating safety, competitive construction, and operating costs, including the streamlining of the licensing process and consumer confidence in nuclear power technology. Thus, during the past decade, much private and government-sponsored research has been directed toward nuclear reactor research in three areas, namely, light water reactors (LWRs), high-temperature gas-cooled reactors (HTGRs), and liquid metal-cooled reactors (LMRs).

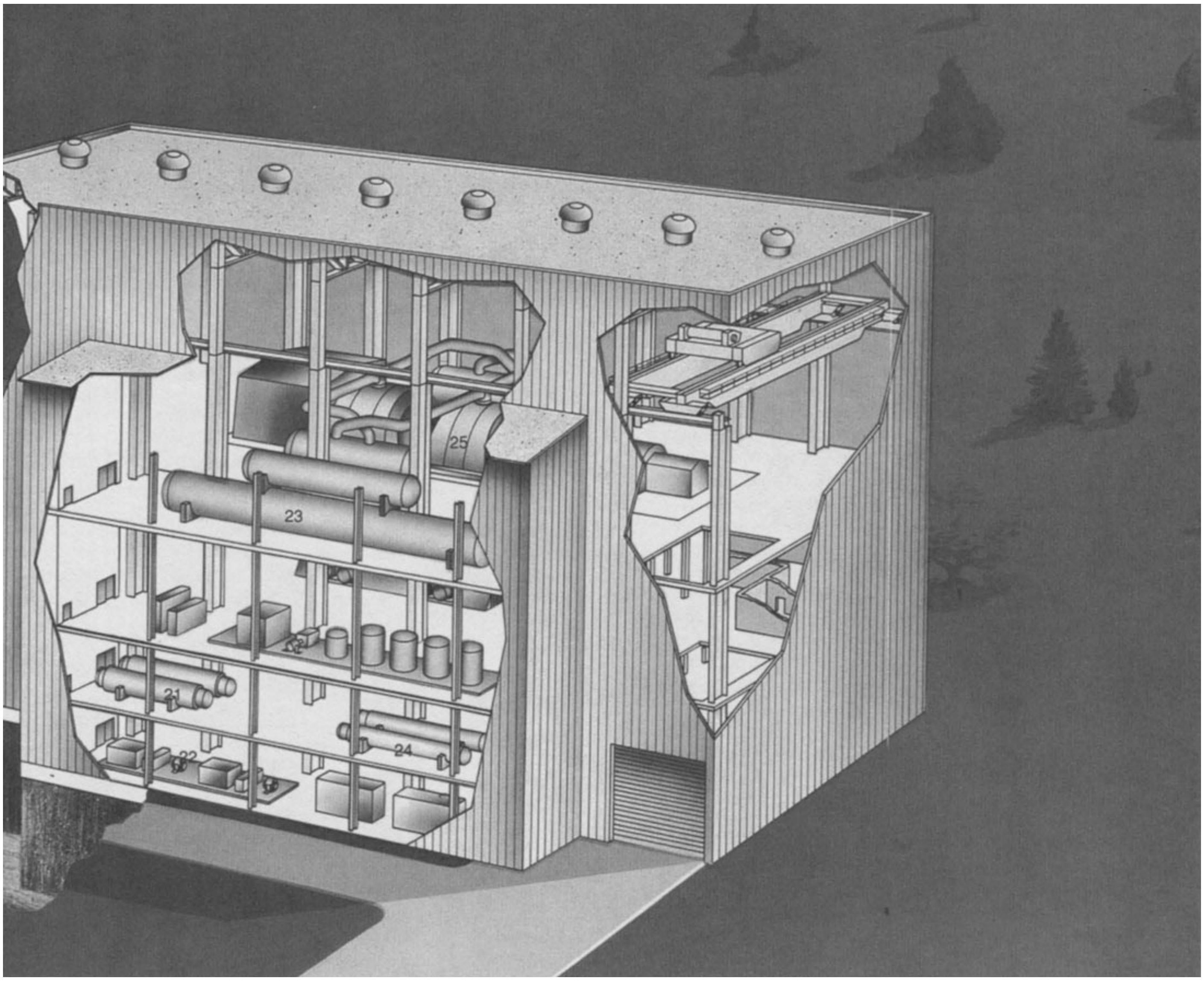


Fig. 35. (*continued*)

21. High Pressure Feedwater Heaters
22. Feedwater Pumps
23. Deaerator
24. Low Pressure Feedwater Heaters
25. Turbine/Generator
26. Spargers (2)
27. Accumulators (2)
28. Main Steam Line
29. Feedwater Line
30. Passive Containment Cooling Air Flow

NUCLEAR POWER INNOVATIONS FOR 1995 AND BEYOND

Increasing concerns over the impact of fossil fuel-burning electric power generating plants on the environment, the accelerating demands for electric power, and a growing dependence on foreign oil supplies brought about a resurgence of interest in nuclear power reactor research

The innovative program commenced with an impressive nuclear power base—some 107 nuclear plants with full-power licenses operating in the United States and producing 18% of the nation's electricity—worldwide, 414 plants in 26 countries generate 298,000 megawatts of electricity (MWe), accounting for 16% of the world's generating capacity.

A survey of reactor developments of the three aforementioned types reveals a number of common generic technical features. These include passive stability, simplification, ruggedness, ease of operation, and modularity.

Overall goals for the innovative program can be summarized as follows:

1. Assured safety with features that minimize the negative consequences of human error, especially a reduction in the chance of occurrence of severe core damage by at least a factor of 10 less than former, contemporary designs.
2. Significantly simpler designs, with increased safety and performance margins in key operational parameters.
3. High reliability throughout a lifetime on the order of 60 years and an increase in plant availability to 85% or greater than the contemporary average of less than 70%.
4. Reduction in capital, operating, maintenance, and fuel costs to meet the economic competition with coal-burning generators. A reduction in construction time to the range of 3 to 5 years as compared with more than 10 years, which has been the experience with some of the later contemporary reactors.
5. A modular design that is standardized at a high-quality level and thus predictably licensable.

Passive Stability. Passive design characteristics ensure core stability by eliminating the potential for a runaway chain reaction. In the innovative program, this has been a hallmark from the outset of the program. Passive characteristics are internal governors—that is, physical laws ensure that the reaction rate decreases instantaneously as the temperature of the coolant or fuel or the power of the reactor increases, without the need for external control devices.

Ruggedness. In some past designs, long-term reliability has been impaired by attempts to achieve the highest in efficiency and economic performance. In response to this negative experience, the margin in certain key performance parameters is being increased in order to lessen the burden on the equipment. By reducing power densities and coolant temperatures, higher reliability will be achieved over a longer lifetime. Past field experience has identified more effective methods of coolant chemistry control and materials selection, factors that will contribute to the long-lived reliability of the components of future systems. Greater emphasis is being given to the selection of proven, high-quality materials and components and on improved methods and quality control over assembly and construction.

Ease of Operation. Thorough investigation of the Three Mile Island incident some years ago showed that a lack of attention had been given to the human factor. The innovation program is addressing this problem in several ways. The computer and telecommunication revolution has made it more practical to use improved technology and human engineering methodologies to revamp the control room and the reactor instrumentation system. These improvements will make the plant easier to operate and provide the operator with a greatly increased amount and quality of information on plant conditions. Graphic displays, diagnostic aids, and expert systems are being developed for such advanced control rooms.

The other design goals complement the new technology to make the operator's task even easier. The passive safety features substantially extend the response time required of the operators in an emergency condition. The margin being built into the systems provides broader normal operating regimes and longer response times for operator action. Greater emphasis is being placed on simplification of operating procedures.

Modularity. Economic competitiveness requires that the construction time be shortened dramatically. Modular construction techniques are a key contributor to achieving this goal and are a proven approach to cost control in major construction projects. Modularization provides for a larger percentage of factory construction, rather than field construction. New innovative concepts will rely heavily on modularization and will be centered around lower unit power outputs, factory assembly, and transportation of modules to the plant site. The overall plant size target is 600 MWe net electrical output.

The AP600 Advanced Plant

As of 1994, the AP600 (Westinghouse) plant is in the most advanced stage of the innovative program. The first unit may be put on line just prior to the expiration of this century. This design satisfies all of the previously mentioned goals of the innovative program. A sectional view of the AP600 is shown in Fig. 35. The site plan is shown in Fig. 36.

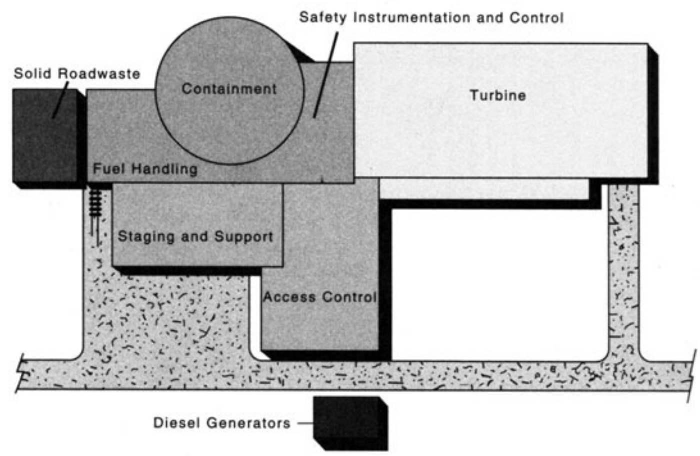


Fig. 36. Site plan for the advanced pressurized water reactor (PWR). (*Westinghouse Electric Corporation, Energy Systems.*)

AP600 Passive Safety System Details. These features reduce operator responsibilities and add an extra margin of safety over contemporary PWR designs. See Fig. 37. Large volumes of water stored in the containment eliminate the need for operator action to assure make-up water, either for small leaks that may occur during normal operation or for a major loss of coolant accident (LOCA). A passive plant is a system that assures public safety even if the operators fail to act.

The passive residual heat removal heat exchangers remove core decay heat if steam generator heat removal is not available. Passive residual heat removal heat exchangers in the in-containment refueling water storage tanks are connected to the reactor coolant system (RCS)

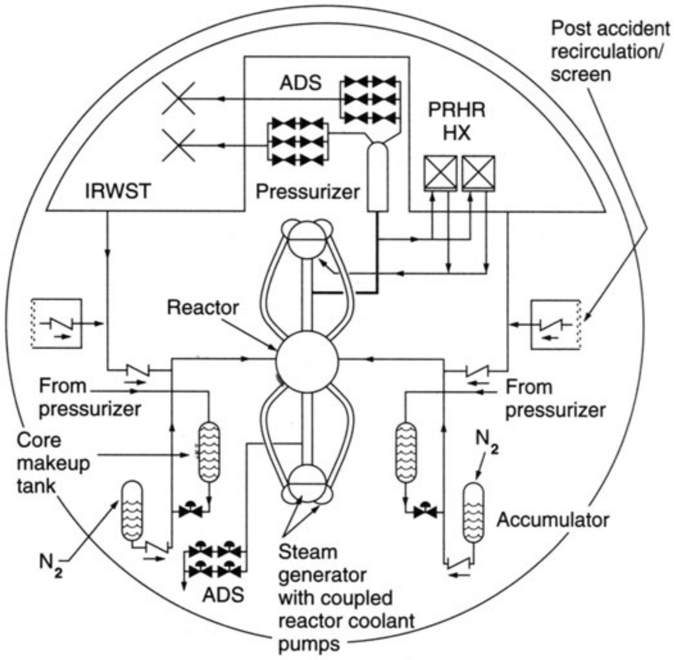


Fig. 37. Schematic representation of the in-containment passive safety injection system (PSIS). IRWST = in-containment refueling water storage tank. PRHR-HX = passive residual heat removal heat exchanger. ADS = automatic depressurization system (four stages). (*Westinghouse.*)

piping, forming a full-pressure, closed, natural circulation cooling loop.

Two core make-up tanks provide borated make-up water whenever the normal make-up system is unavailable. The tanks are located above the reactor coolant system loop piping and kept at system pressure by steam lines from the pressurizer. These tanks function at any system pressure, using only gravity as a motive force. If the reactor protection system detects a need for make-up water, core make-up tanks discharge and isolation valves open automatically, allowing the tanks to drain into the reactor vessel.

Two accumulators provide the high make-up flows initially required by a large LOCA. These tanks contain 1700 cubic feet (about 48 cubic meters) of borated water pressurized with 300 cubic feet (about 8.5 cubic meters) of nitrogen at 700 psi (4.8×10^6 Pa). The accumulators are isolated from the RCS by check valves. Each accumulator is paired with a core make-up tank, the pair sharing an injection line to the reactor vessel downcomer. This ensures that at least one accumulator/core make-up tank pair would be available following an injection line LOCA.

The in-containment refueling water storage tank provides 500,000 gallons (about 1900 cubic meters) of water with a gravity head above the core. This water inventory is sufficient to flood the containment above the level of the reactor core and provide decay heat removal by natural circulation.

The automatic depressurization system depressurizes the RCS if core make-up tank level is low. Depressurization allows gravity injection from the in-containment refueling water storage tank, which is at atmospheric pressure. To ensure that the automatic depressurization system works when needed, while minimizing the consequences of spurious valve operation, the system provides phased depressurization with two redundant sets of valves connected to the pressurizer. The discharge is sparged into the in-containment refueling water storage tank. The automatic depressurization system valves are arranged in three stages to reduce peak flow rates. A fourth depressurization stage is provided directly on the RCS hot leg.

The passive containment cooling system provides the safety grade ultimate heat sink that prevents the containment shell from exceeding its design pressure of 45 psig [3.1×10^5 Pa (g)]. The system uses natural air circulation between the steel containment shell and the concrete shield building. During postulated accidents, air cooling is enhanced by draining water onto the steel containment shell. The water is provided by gravity from a 350,000-gallon (about 1300 cubic meters) annular tank in the roof of the shield building. This tank has sufficient water to provide three days of cooling.

RADIOACTIVE WASTES

Decisions pertaining to the location of radioactive waste sites mainly derive from political and sociological sources. A condensed overview of the technical aspects is given here.

The radioactive wastes associated with nuclear reactors fall into two categories: (1) *commercial wastes*—the result of operating nuclear-powered electric generating facilities; and (2) *military wastes*—the result of reactor operations associated with weapons manufacture. Because the fuel in plutonium production reactors, as required by weapons, is irradiated less than the fuel in commercial power reactors, the military wastes contain fewer fission products and thus are not as active radiologically or thermally. They are nevertheless hazardous and require careful disposal.

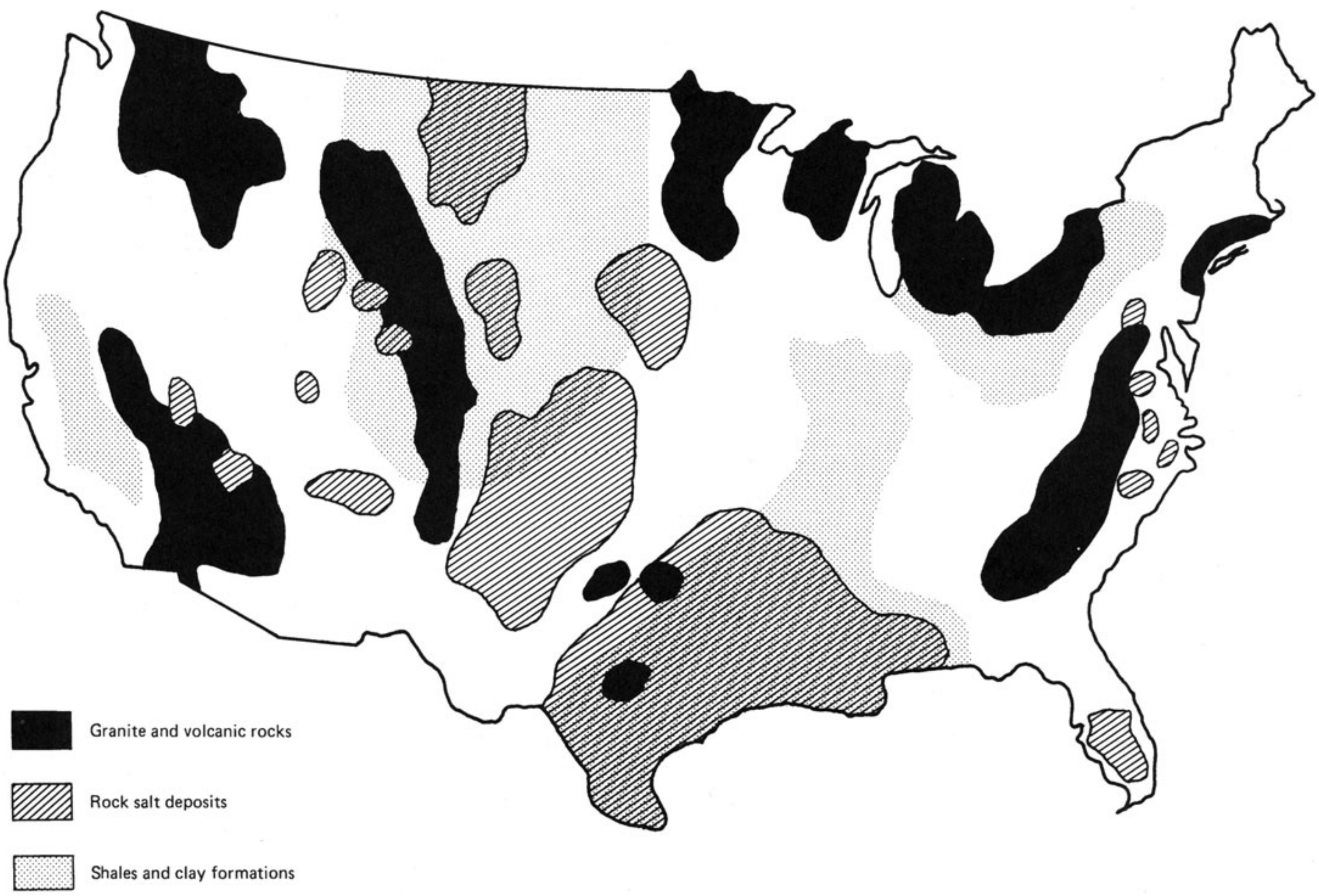


Fig. 38. Some authorities prefer rock salt deposits for the permanent disposal of nuclear wastes on the assumption that the heat generated by radioactive decay would fuse salt and wastes into an impermeable mass. Other experts question the integrity of salt formations. Increasing attention has turned to hard media, such as the granitic and basaltic rocks, and to shale and clay formation, with the hope that the extensive occurrence of such formations would minimize the need to transport wastes over long distances.

Nuclear power plants use fuel rods with a life span of about three years. Each year, roughly one-third of spent fuel rods are removed and stored in cooling basins, either at the reactor site or elsewhere. Typical modern nuclear power plants discharge about 30 tons of the spent fuel per reactor per year. Comparatively little of the radioactive wastes, as is currently reliably known worldwide, has been processed for return to the fuel cycle. Actually, fuel reprocessing causes a net increase in the volume of radioactive wastes, but, as in the case of military wastes, they are less hazardous in the long term. Nevertheless, the wastes from reprocessing also must be disposed of with great care.

Spent fuel from a reactor contains unused uranium as well as plutonium-239 which has been created by bombardment of neutrons during the fission process. Mixed with these useful materials are other highly radioactive and hazardous fission products, such as cesium-137 and strontium-90. Since reprocessed fuels contain plutonium, well suited for making nuclear weapons, concern has been expressed over the possible capture of some of this material by agents or terrorists operating on behalf of unfriendly governments that do not have a nuclear weapons capability.

Categories of Wastes by Content

In addition to the two source categories previously mentioned, radioactive wastes are classified in accordance with their content:

High-level wastes contain 99.9% of the nonvolatile fission products, 0.5% of the uranium and plutonium, and all the actinides formed by transmutation of the uranium and plutonium in the reactors. Among the actinides are neptunium and americium. High-level wastes are either the aqueous wastes resulting from reprocessing; or the spent-fuel rods to be disposed of in the absence of reprocessing.

Cladding wastes are comprised of solid fragments of Zircaloy and stainless steel cladding (tube in which the fuel is placed) and other structural elements of the fuel assemblies remaining after the final cores have been dissolved.

Low-level transuranic wastes are solid or solidified materials which contain plutonium or other long-lived alpha-particle emitters in known or suspected concentrations higher than 10 nanoCuries per gram and external radiation levels after packaging sufficiently low to allow direct handling.

Intermediate level transuranic wastes are solids or solidified materials that contain long-lived alpha-particle emitters at concentrations greater than 10 nanoCuries per gram and which have, after packaging, typical surface dose rates between 10 and 1000 mrems/hour due to fission product contamination.

Nontransuranic low-level wastes are diverse materials which are contaminated with low levels of beta- and gamma-emitting isotopes,

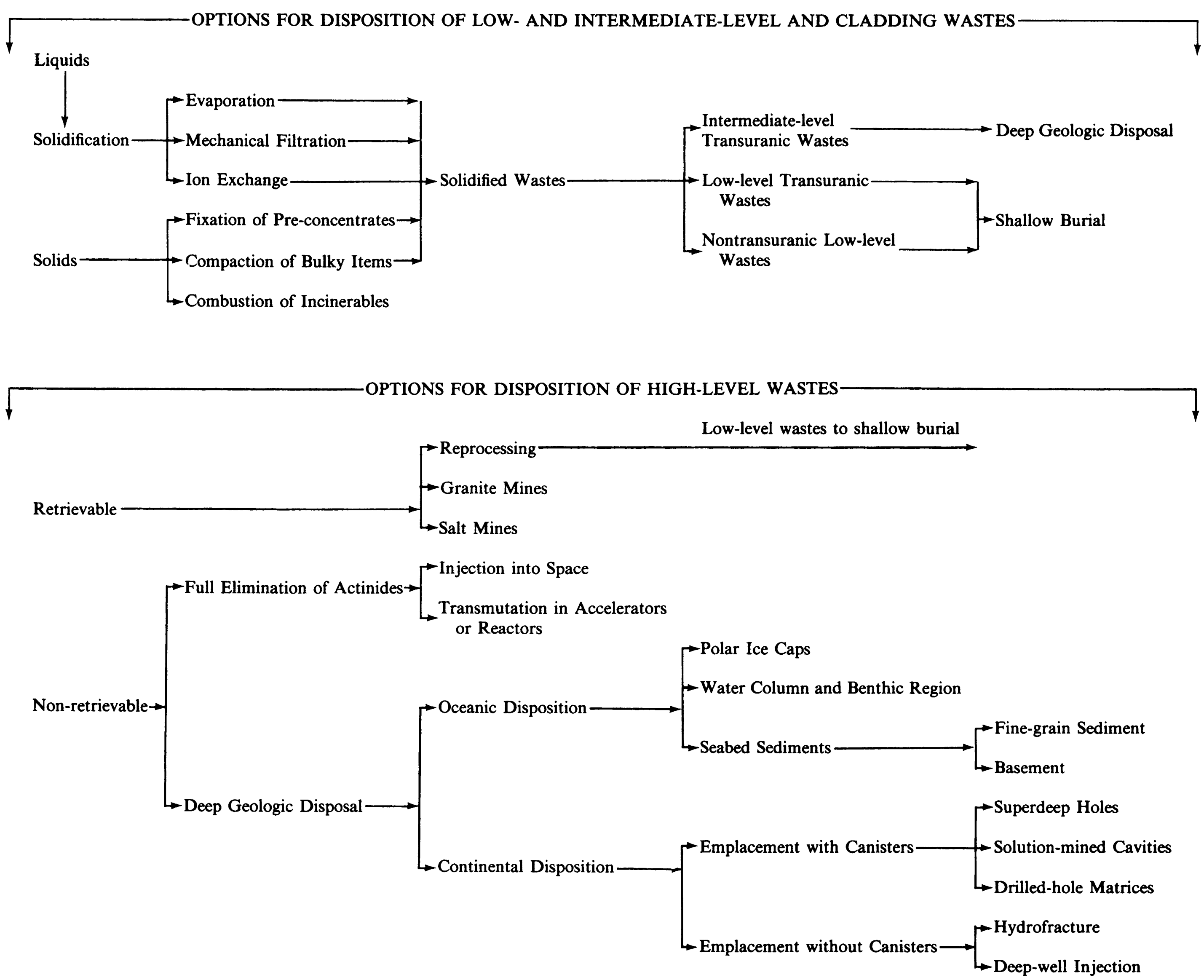


Fig. 39. Panorama of options for consideration in the disposition of radioactive wastes. (*Top portion of chart*: Options for disposition of low- and intermediate-level and cladding wastes. *Bottom portion of chart*: Options for disposition of high-level wastes).

but which contain less than 10 nanoCuries of long-lived alpha activity per gram.

Permanent Disposal Methodologies

Because of many doubts among various authorities pertaining to the permanent geologic depositories, as previously mentioned, considerable effort continues as regards semipermanent and permanent types of depositories. Many methodologies involve the so-called "sequence of barriers" approach. The first barrier is the form in which radioactive materials are embedded—vitrification, calcination, etc. The requirements for the first barrier are that it not be corrosive and possess excellent thermal stability and mechanical integrity. Wastes generate much heat during their initial decade of confinement. This affects decisions as regards the wasteform and the second barrier, the frequently mentioned canister which encapsulates the wasteform. The principal function served by the canister is protection of the material during the collection and transportation (to geologic site) phases. The canisters also should provide excellent protection of their contents for a minimum of 50 years, just in case it is desired to retrieve the wastes at some future date. Canisters must resist corrosive chemicals, they must withstand extremely high radiation fluxes caused by fission-product decay and the heat generated by the decaying wastes. It is interesting to note that an unprotected stainless steel canister will not resist structural deterioration arising from salt brines for that long a period. Provision must be made for cooling the canisters, either by air or water. The canisters should be designed to permit maximum heat transfer and, currently, the cylinder and annulus configurations are preferred. A third barrier would be the geologic site itself, obviously impervious to water penetration and in a seismically stable location. To fulfill all the foregoing requirements (and more), consideration is being given to phasing the waste storage procedure, possibly storing the canisters for water cooling during the first few years, after which air cooling would suffice.

The predominant preference appears to be the underground depository option. Along these lines, it is interesting to review the rock formations in the United States, as shown in Fig. 38.

Commenting briefly on other proposed methodologies, as shown in Fig. 39: (a) In *solution-mined cavities*, it is proposed that chemical solutions would be used to mine cavities in appropriate media, such as rock salt; (b) in the *drilled-hole matrix*. A series of large-diameter holes would be drilled into the geologic media to depths up to 2 kilometers to form a grid of holes. The solid wastes would be packed into these holes, then sealed. (c) In the *rock-melting concept*, liquid wastes (no solidification) are poured into a subterranean cavity which would be created by an underground explosion. (d) In the *hydrofracture concept*, liquid radioactive wastes are converted into a type of grout (cement or cementlike materials used). This grout is pumped under high pressure into shale as deep as 1 kilometer. The pressure of the operation causes the underlying shale to fracture and the wastes fill up the cracks so formed. This procedure has been used for years in the petroleum field. See also **Petroleum**. (e) In the *polar ice* concept, the wastes would melt through the ice (although this approach would require considerable new technology); or the wastes would be placed on the surface of the ice or anchored within the ice. Advantages include long distances from populations and excellent thermal cooling. Disadvantages include extensive transport and poor retrievability. This method is not high on the list of choices mainly because of too many unknown factors that will require considerable research and experimentation. (f) *Oceanic disposition*, in addition to the polar ice cap concept, are subduction zones and other deep sea trenches and rapid sedimentation areas. A "Seabed Disposal Program Annual Report" states, "Placing high-level wastes on the seafloor, i.e., in the water column, effectively puts the waste contained directly into the biosphere. Since it is difficult to conceive of a practical man-made wasteform/container system that would survive without releasing radionuclides for hundreds of thousands of years in a marine environment, one must assume the radioactive material would eventually enter the ecosystem." The sub-seabed sediments in the central North Pacific are loosely packed, fine-grained, deepsea "red clays" and have not been fully dismissed from continuing investigation.

Additional Reading

Apostolakis, G.: "The Concept of Probability in Safety Assessments of Technological Systems," *Science*, 1359 (December 7, 1990).

Bothwell, R.: "Nucleus. The History of Atomic Energy Limited," Univ. of Toronto, Toronto, Canada, 1988.

Bruschi, H., and T. Andersen: "Turning the Key," *Nuclear Engineering International*, 1 (November 1991).

Cobb, C. E., Jr., and K. Kasmauski: "Living with Radiation," *National Geographic*, 403 (April 1989).

Cottrell, A.: "Safe Energy," *Review (University of Wales)*, 38 (Autumn 1987).

Davis, W. K.: "Problems and Prospects for Nuclear Power," *Chem. Eng. Progress*, **80**(6), 11–16 (June 1984).

Golay, M. W., and N. E. Todreas: "Advanced Light-Water Reactors," *Sci. Amer.*, 82 (April 1990).

Golay, M. W.: "Longer Life for Nuclear Plants," *Technology Review (MIT.)*, 25 (May/June 1990).

Goldschmidt, B.: "Atomic Rivals," Rutgers Univ. Press, New Brunswick, New Jersey, 1990.

Green, S. J.: "Solving Chemical and Mechanical Problems of PWR Steam Generators," *Chem. Eng. Progress*, 31 (July 1987).

Hansen, K., et al.: "Making Nuclear Power Work: Lessons from Around the World," *Technology Rev. (MIT)*, 30 (February 1989).

Jerome, F.: "Yo-Yo Journalism and Nuclear Power," *Technology Review (MIT)*, 73 (April 1989).

Kairi, S. P.: "Outage Risk Management," *EPRI J.*, 34 (April/May 1992).

Lester, R. K.: "Rethinking Nuclear Power," *Sci. Amer.*, 31 (March 1986).

Marshall, E.: "Counting on New Nukes," *Science*, 1024 (March 2, 1990).

Miller, P., and R. H. Ressmeyer: "A Comeback for Nuclear Power? Our Electric Future," *National Geographic*, 60 (August 1991).

Roberts, L.: "British Radiation Study Throws Experts into Tizzy," *Science*, 24 (April 6, 1990).

Shoup, R. L.: "International Waste Management Symposium," *Nuclear Safety*, **18**, 4, (1977).

Skerret, P. J.: "Will the Public Say Yes to Nukes?" *Technology Review MIT*, 8 (April 1991).

Slovic, P., Flynn, J. H., and M. Layman: "Perceived Risk, Trust, and the Politics of Nuclear Waste," *Science*, 1603 (December 13, 1991).

Spinard, B. I.: "U.S. Nuclear Power in the Next Twenty Years," *Science*, 707 (December 12, 1988).

Staff: "AP600—A Cost Competitive Power Source," *Energy Digest*, 4 (Pittsburgh, Pennsylvania (Fall 1991).

Suzuki, T.: "Japan's Nuclear Dilemma," *Technology Review (MIT)*, 41 (October 1991).

Taylor, J. J.: "Improved and Safer Nuclear Power," *Science*, 318 (April 21, 1989).

Weinberg, A. M., and I. Spiewak: "Inherently Safe Reactors and a Second Nuclear Era," *Science*, **224**, 1398–1402 (1984).

NUCLEAR SPIN. The intrinsic angular momentum of the atomic nucleus due to rotation about its own axis. It is usually designated I and has the magnitude, $\sqrt{I(I+1)}\,h/2\pi \approx I(h/2\pi)$, where I is the nuclear spin quantum number which has different (integral or half-integral) values (including zero) for different nuclei. In spectroscopy, the nuclear spin is of importance for the explanation of the hyperfine structure, and of the intensity alternation in band spectra.

NUCLEAR STRUCTURE. The nucleus of an atom of atomic number Z and mass number A contains Z protons and A-Z neutrons, bound together under the influence of shortrange nuclear forces much as molecules are bound together in a drop of liquid. The strength of binding may be determined by subtracting the actual mass of the atom from the mass of its constitutent particles considered as free particles. The binding energy E_B is then related to this mass defect ΔM by Einstein's relation, $E_B = \Delta Mc^2$, where c is the velocity of light. The precise value of E_B depends upon the nucleus concerned, and upon how many neutrons and protons it contains but it is of the order of 8 MeV per nucleon in most nuclei.

The binding energy determines whether the nucleus is stable or unstable. Among the lighter nuclei, the ones which are stable are those in which the number of protons is approximately equal to the number of neutrons, so that $A \approx 2Z$. In heavier stable nuclei, there is an excess of neutrons over protons owing to the repulsive electrostatic forces between the protons. Thus, the most stable oxygen nucleus is ^{16}O, containing 8 protons and 8 neutrons, while the most nearly stable uranium nucleus is ^{238}U, containing 92 protons and 146 neutrons. Nuclei containing a disproportionate number of neutrons tend to be unstable and decay radioactively by emission of electrons whereby neutrons are converted into protons; those containing an excess of protons similarly tend to decay by emission of positrons or by capture of orbital electrons.

It has been shown that the nucleus is approximately spherical in shape and of volume proportional approximately to its mass. It is, however, capable of executing oscillations about the spherical form, and in certain circumstances may even acquire a permanent deformation. The heaviest nuclei are unstable under deformation, as a result of which they undergo spontaneous fission. These properties may be described qualitatively by regarding the nucleus as an electrically charged drop of liquid possessing volume energy and surface tension.

Although the nucleus is normally found in its lowest energy state, it may be produced as the result of a nuclear reaction, or through radioactivity in a number of excited states whose detailed properties may differ quite markedly from the lowest state. If formed in an excited state, it will decay, normally by the emission of electromagnetic radiation (gamma rays) to the lowest state, or by the emission of particles to another nucleus.

NUCLEAR TRANSMUTATION. The transformation of one nuclide into another, which differs from it in nuclear charge, mass, or stability, i.e., in a nuclear reaction or a process of radioactivity. Such changes occur in natural radioactive processes, but the general need for a systematic notation for expressing them came only with the great number of transmutations discovered after particle accelerators provided high-energy ion beams capable of penetrating the Coulomb barrier of all stable atomic nuclei.

Two representative transmutation equations are

$$^{27}\mathrm{Al} + n \rightarrow {}^{27}\mathrm{Mg} + {}^{1}\mathrm{H}$$

which shows the transmutation of aluminum atoms of mass number 27, by bombardment with neutrons, to magnesium atoms of mass number 27, with the emission of a proton, and

$$^{9}\mathrm{Be} + \gamma \rightarrow {}^{8}\mathrm{Be} + n$$

which shows the transmutation of beryllium atoms of mass number 9, under gamma-ray bombardment, to beryllium atoms of mass number 8, with the emission of a neutron.

The two reactions above may also be expressed in condensed from as

$$^{27}\mathrm{Al}(n, p)^{27}\mathrm{Mg}$$

and

$$^{9}\mathrm{Be}(\gamma, n)^{8}\mathrm{Be}.$$

NUCLEATE BOILING. See **Boiler; Boiling.**

NUCLEIC ACIDS AND NUCLEOPROTEINS. Nucleic acids are compounds in which phosphoric acid is combined with carbohydrates and with bases derived from purine and pyrimidine. Nucleoproteins are conjugated proteins consisting of a protein moiety and a nucleic acid. Originally, nucleoproteins were thought to occur only in the nuclei of cells, but it was later established that they are far more widely distributed, being found in cells of all types, animal and plant. They are found in the chromosomes, in the genes, in viruses, and bacteriophages.

The protein portion of the nucleoproteins is basic in nature and being complex in structure may form several types of linkage, depending upon the type of nucleic acid. In gastric digestion or hydrolysis with weak acid, nucleoproteins yield protein and nuclein. The latter in pancreatic digestion or hydrolysis with weak alkali yields additional protein and nucleic acid.

Upon additional hydrolysis, nucleic acids yield four characteristic constituent groups: (1) heterocyclic nitrogenous bases of the purine type; (2) heterocyclic nitrogenous bases of the pyrimidine type; (3) a carbohydrate, either ribose or deoxiribose; and (4) phosphoric acid.

Hydrolysis of nucleic acid with enzymes belonging to the group of nucleases gives nucleotides, nucleosides, and the constituent groups already mentioned. Thus, polynucleotidase catalyzes the hydrolysis of nucleic acid to give nucleotides which consist of purine or pyrimidine, ribose or deoxyribose, and phosphoric acid. Nucleotidase catalyzes the hydrolysis of nucleotides to nucleosides (which consist of a purine or a pyrimidine and ribose or deoxyribose) and phosphoric acid. Nucleosidase catalyzes the acid hydrolysis of nucleosides to the respective base or carbohydrate. Since the nucleic acids are composed of nucleotides, they may be considered polymers of nucleotides and thus be called polynucleotides.

There are two groups of nucleic acids differentiated by the carbohydrate present. Those which contain *D*-ribose,

$$\mathrm{CHO{\cdot}CHOH{\cdot}CHOH{\cdot}CHOH{\cdot}CH_2OH},$$

are generally known as *ribonucleic acids*, usually termed RNA. Those which contain D-2-deoxyribose,

$$\mathrm{CHO{\cdot}CH_2{\cdot}CHOH{\cdot}CHOH{\cdot}CH_2OH},$$

are usually known as *deoxyribonucleic acids*, generally termed DNA. Sometimes the latter are also termed deoxyribonucleic acid or DRNA.

Deoxyribonucleic acid is found not only in the nucleus of normal cells but is also present in mitochondria of both plants and animals as well as in plant chloroplasts. DNA occurs in circular, double standard structures. Formerly, this type of nucleic acid was thought to be a constituent only of animal cells and was known as thymus nucleic acid. Ribonucleic acid is found principally in the cytoplasm, although small amounts are found the nucleus, nucleoli, and the chromosomes. RNA at one time was known as yeast nucleic acid and was thought to be the characteristic nucleic acid of plants.

Deoxyribonucleic acid is a constituent of the chromosomes of the cell nucleus. Since the number of chromosomes of a cell and its daughter cells are equal, the quantity of DNA in the normal cells of any given species or type should be and is remarkably constant. This quantity is not changed by starvation or other action or form of stress. The quantity in a normal diploid nucleus is twice that of a normal haploid nucleus and in polyploid cells, such as cultivated wheats, the quantity of DNA is some multiple of the quantity in the haploid cell. This quantity is of the order of 6×10^{-9} milligrams per nucleus in the case of mammals. Birds and fishes have lesser amounts.

At one time, the chemical synthesis of DNA and the resultant ability to construct totally synthetic genes appeared to be impractical. However, under the influence of recombinant DNA technology, the field of synthetic DNA has matured. Two chemical methods—diester and triester—and one enzymatic method—polynucleotide phosphorylase—are generally used for the synthesis of oligodeoxyribonucleotide. Once made the resultant gene can be cloned (see **Clone**) and then used to produce useful peptide products. Products such as insulin and somatostatin have been made in *Escherichia coli* following insertion of the chemically synthesized DNA into the bacterial gene.

The quantity of RNA varies in different tissues. It also varies in amount in the cytoplasm and nucleus of the same cells. The amount of RNA is affected by the nutritional state of the cells, the type of tissue, and the metabolic action in which they participate. Thus, there is more ribonucleic acid in growing embryonic tissue, in tumor tissue, in pancreas, salivary glands, and other gland with secretory functions. All of these indications led to the concept that RNA controls the rate of the production of protein by cells, while DNA controls the transmission of hereditary characteristics from one generation to the next.

At one time, it was believed that nucleic acids were tetranucleotides, but it has been established that they are much larger molecules. Ribonucleic acids have been found to have molecular weights as high as 300,000 (implying approximately 1,000 nucleotides per molecule), while deoxyribonucleic acids have molecular weights of the order of 1 or 2 million. It has been shown that in the living organism, RNA and DNA are long double chains, with each chain spiraling around the other, also a characteristic of some proteins.

Each of the two spiral chains that constitute the DNA molecule is composed of alternating sugar (deoxyribose or ribose) and phosphate groups, and attached to each sugar is one of the four bases, usually adenine, guanine, thymine, and cytosine. (Adenine and guanine are pyrimidines, while thymine and cytosine are purines). Linkage between the two chains is effected by hydrogen bonds formed between the adenine groups of each chain and the thymine groups of the other, and between the guanine groups of each chain and the cytosine groups of the other. Variations occur in the sequence in which these four bases appear in each chain. These variations are the means of transmission of genetic "information," e.g., a basic mechanism of heredity.

Delineation of DNA as the bearer of genetic information stems from the findings of Chargaff et al. that the base composition of DNA is related to the species of origin. Indeed Chargaff's chromatographic studies enabled the establishment of the following conclusions:

1. The base composition of DNA varies from one species to another.
2. DNA specimens isolated from different tissues of the same species have the same base composition.
3. The base composition of DNA in a given species does not change with age, nutritional status, or changes in environment.
4. In nearly all DNAs examined, the number of adenine residues is always equal to the number of thymine residues, i.e., A = T, and the number of guanine residues is always equal to the number of cytosine residues, G = C. As a corollary, it is clear that the sum of purine residues equals the sum of pyrimidine residues, i.e., A + G = C + T.
5. The DNAs extracted from closely related species have similar base composition, whereas those of widely different species are likely to have widely different base composition.

See also **Adenosine; Adenosine Phosphates; Amino Acids; Bacteriophage; Cell (Biology); Genes and Genetics; Gene Science; Heredity;** and **Proteins.**

Ann C. Vickery, Ph.D., Associate Professor,
College of Public Health, University of South Florida,
Tampa, Florida.

NUCLEI (Cloud Formation). See **Atmosphere-Ocean Interface.**

NUCLEI (Sublimiation). See **Precipitation and Hydrometeors.**

NUCLEONICS. The applications of nuclear science in physics, chemistry, biology, and other sciences, including military science, and in industry, and the techniques associated with these applications.

NUCLEONS. Two nuclear particles, the proton and the neutron, and their antiparticles are known as *nucleons*. The rest mass of the proton is 1.0076 amu; that of the neutron, 1.0089 amu. The antiproton bears the same relation to the proton that the electron does to the positron, i.e., its charge is equal and opposite and its mass is the same, the charge being equal in magnitude to the electronic charge. Protons and antiprotons also annihilate each other when they collide, the reaction of a single pair producing positive and negative pions or kaons. If the proton and antiproton do not collide, but experience a "near miss," then an exchange of charge can occur, resulting in the formation of a neutron-anti-neutron pair. The four nucleon particles are fermions and have a spin angular momentum quantum number of $\frac{1}{2}$. See also **Neutron; Particles (Subatomic);** and **Proton.**

NUCLEOPHILE. An ion or molecule that donates a pair of electrons to an atomic nucleus to form a covalent bond. The nucleus that accepts the electrons is called an *electrophile*. This occurs, for example, in the formation of acids and bases according to the Lewis concept, as well as in covalent bonding in organic compounds.

NUCLEOPHILIC REACTION. A reaction in which a nucleophilic reagent attacks an electrophilic 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 reagent.

NUCLEUS. 1. The nucleus of an atom is the positively-charged core, with which is associated practically the entire mass of the atom, but only a minute part of its volume.

2. The nucleus of a molecule is a group of atoms connected by valence bonds so that the atoms and their bonds form a ring or closed structure, which persists as a unit through a series of chemical changes.

3. A group of cell bodies in the central nervous system of vertebrates. Examples of such groups are the red nucleus in the midbrain, through which impulses are routed for the control of subconscious muscular movements, and Deiter's nucleus, lying at the junction of the medulla with the hindbrain. Through this center impulses pass for muscular action involved in the maintenance of equilibrium.

4. A somewhat spherical or oblong body in most living cells. This nucleus contains the chromosomes which, in turn, bear the genes of heredity. The nucleus also contains a nucleolus, or sometimes two or more nucleoli and a basic ground substance, the nucleoplasm. A nuclear membrane surrounds it on the outside, but this membrane is very porous, allowing materials to pass through rather freely.

5. Condensation nuclei take part in phase changes, as in the formation of clouds; in seeding concentrated solutions to bring about crystallization, precipitation, etc.

NUCLIDES. See **Chemical Elements.**

NULL VECTOR. A vector $\mathbf{A}_\mu$ of zero length ($\mathbf{A}_\mu \mathbf{A}_\mu = 0$). In special relativity theory the displacement between two events on the path of a photon is a null vector.

NUMBER THEORY. The theory of numbers is concerned primarily with the properties of the integers, or whole numbers, $0, \pm 1, \pm 2, \ldots$. The integers form a ring, i.e., when one or more of the operations addition, subtraction, and multiplication are applied to any two integers, the result is always an integer.

Numbers can be classified in many ways. For example, odd and even numbers, prime numbers, square numbers, and perfect numbers. These classes of numbers were mentioned by Euclid (circa 300 B.C.). The Chinese knew in 500 B.C. that $2^p - 2$ is divisible by p, for prime numbers p. Euclid proved that there exist infinitely many primes. The proof is by contradiction. Assume p is the largest prime. Let M be the product $2 \cdot 3 \cdot 5 \cdots p$ of all primes up to p. Then $M + 1$ is either a prime number itself, or is divisible only by primes greater than p.

While most results of number theory are easy to understand, it has taken brilliant mathematicians many years to construct proofs of many and some of the most interesting conjectures remain unproved.

Number theory has been extended to the study of the specific properties of other classes of numbers, such as rational, algebraic, and transcendental numbers. Discussed here are elementary number theory, algebraic number theory, analytic number theory, and Diophantine approximations.

Elementary Number Theory. The main concern of this branch of number theory is divisibility. If a and b are integers, and $a = bc$ for some c, then b is a factor or divisor of a, written symbolically as $b|a$. Thus, $1|a$ and $a|a$ for every integer a. A prime number $p > 1$ is an integer that has only 1 and p as divisors. Other integers greater than 1 are called composite.

A concept known as the fundamental theorem of arithmetic which was proved by Euclid states that every integer n greater than 1 can be expressed as a product of primes in only one way when the order of prime factors is disregarded.

A perfect number is an integer which is equal to the sum of its proper divisors less than itself. The five least are: 6; 28; 496; 8,128; 33,550,336. Euclid showed that $(2^n - 1)2^{n-1}$ is a perfect number if and only if $2^n - 1$ is a prime number. Numbers of the form $2^p - 1$ are known as Mersenne numbers. All even perfect numbers are of Euclid's type. No odd perfect numbers have been found.

Let m denote a positive integer. Every integer a can be represented in the form $a = qm + r$, where q is an integer and r is an integer which has one of the m values $0, 1, \ldots, m - 1$. Two integers a and b are congruent modulo m if the value of r is the same for a and b when expressed in the above form. Symbolically, $a \equiv b(\text{mod } m)$. If $a \equiv b(\text{mod } m)$, then $m|(a - b)$. The congruence relationship is an equivalence, i.e., it is reflexive, symmetric, and transitive. The relation defines equivalence classes of numbers congruent to each other which are called residue classes. There are m residue classes modulo m.

The symbol (a, b) is frequently used to denote the greatest common divisor of integers a and b, assumed at least one of which is nonzero. If $(a, b) = 1$, a and b are coprime.

If $a \equiv b \pmod m$, $(a, m) = (b, m)$. If $(a, m) = 1$, the residue class of a modulo m is said to be prime to m. Those residue classes which are prime to m form a reduced system of residue classes mod m. Euler introduced the function $\Phi(m) = m\Pi_{p/m}(1 - 1/p)$ to denote the number of reduced residue classes modulo m. Some of the properties of Φ are:

$$\Phi(mn) = \Phi(m)\Phi(n) \quad \text{if} \quad (m, n) = 1$$

and

$$\Phi(p^r) = p^r(1 - 1/p)$$

where p is any prime and r is a positive integer.

Two important congruences are Wilson's Theorem

$$(p - 1)! \equiv -1 \pmod p$$

and Fermat's theorem

$$a^{p-1} \equiv 1 \pmod p, \text{ where } p \text{ is a prime.}$$

Euler generalized the latter to

$$a^{\Phi(m)} \equiv 1 \pmod m, \text{ where } (a, m) = 1$$

The linear congruence $ax \equiv b \pmod m$ is solvable for x if and only if $d|b$, where $d = (a, m)$. There are exactly d solutions (incongruent to each other).

If $m = p$ is a prime number then $ax \equiv 1 \pmod p$ has exactly one solution for each a which is coprime with p. This solution is the reciprocal of a modulo p which shows that residue classes modulo p form a finite field.

The congruence $a_0x^n + a_1x^{n-1} + \cdots + a_n \equiv 0 \pmod m$ is of degree n, if m does not divide a_0. If $m = p$, the number of solutions cannot exceed n but a solution may not exist.

If $(a, m) = 1$, the least positive e such that $a^e \equiv \Phi \pmod m$ must divide $\Phi(m)$. If $e = \Phi(m)$, a is called a primitive root of m and the powers of $a, a^2, \ldots, a^{e-1}$ form a reduced set of residues modulo m. A modulus m has primitive roots only when $m = 2, 4, p^k, 2p^k$, where p is an odd prime. The period of the periodic decimal into which the reduced fraction a/m can be expanded has the length $\Phi(m)$ or a divisor thereof. It has the length $\Phi(m)$ only when the base (10 for decimal expansion) is a primitive root modulo m.

Fibonacci numbers were discovered by Leonardo of Pisa in 1202 A.D. The first two Fibonacci numbers are 1; subsequent Fibonacci numbers are obtained by adding the previous two Fibonacci numbers. The first seven are 1, 1, 2, 3, 5, 8, and 13. The sequence grows exponentially with the nth number in the sequence being approximately equal to $(1/\sqrt{5})((1 + \sqrt{5})/2)^n$.

A theorem which was known to the Chinese in ancient times and is known as the Chinese remainder theorem can be stated as:

If $(m_i, m_j) = 1$ for $1 \le i < j \le n$, then the system of linear congruences

$$\begin{aligned} x &\equiv a_1 \pmod{m_1} \\ x &\equiv a_2 \pmod{m_2} \\ &\cdots \\ x &\equiv a_n \pmod{m_n} \end{aligned}$$

has a unique solution modulo m, where $m = \Pi_{i=1}^n m_i$.

Gödel showed how to use this theorem as a coding trick, in which an arbitrary finite sequence of numbers can be encoded as a single number.

In 1970, Yuri Matyasevich completed the proof that no procedure can be devised to determine the existence of a solution to a Diophantine equation. This proof involves the Chinese remainder theorem, Fibonacci numbers, and Diophantine equations. In 1900, David Hilbert posed 23 major problems for mathematicians. Matyasevich's results complete the solution of Hilbert's tenth problem.

If $x^2 \equiv a \pmod m$ is solvable, a is called a quadratic residue modulo m, otherwise a quadratic nonresidue. Let $m = p$ be an odd prime. The Legendre symbol (a/p) is defined to be $+1$, if a is a quadratic residue modulo p and as -1 if a is a quadratic nonresidue. It is assumed that $a \not\equiv 0 \pmod p$.

The Legendre symbol is a character of the multiplicative group of residue classes modulo p:

$$\left(\frac{a}{p}\right) \cdot \left(\frac{b}{p}\right) = \left(\frac{ab}{p}\right)$$

If q is another odd prime

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{(p-1)/2\,(q-1)/2}$$

which is known as the reciprocity law of quadratic residues. Also,

$$\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2}$$

Jacobi generalized the Legendre symbol by defining

$$\left(\frac{a}{p_1p_2\cdots p_n}\right) = \left(\frac{a}{p_1}\right)\left(\frac{a}{p_2}\right)\cdots\left(\frac{a}{p_n}\right), \quad \text{and} \quad \left(\frac{a}{-k}\right) = \left(\frac{a}{k}\right),$$

if a is positive.

The reciprocity law still holds for any odd numbers p and q not both negative.

An expression, $f(x, y) = ax^2 + bxy + cy^2$, is a binary quadratic form. A number m is represented by the form f if there exists a pair of integers u and v such that $f(u, v) = m$. The representation is called primitive if $(u, v) = 1$. The determinant of the form f *is defined as* $d = ac - (\frac{1}{4})b^2$. The discriminant $D = -4d$.

The forms, $f = ax^2 + bxy + cy^2$ and $F + AX^2 + BXY + CY^2$, are equivalent if there exists a transformation $x = \alpha x + \beta y$, $Y = \gamma x + \delta y$ with $\alpha\delta - \beta\gamma = \pm 1$, α, β, δ all integers so that $F(X, Y) = f(x, y)$.

The determinants of equivalent forms are equal.

Since $af = (ax + \frac{1}{2}by)^2 + dy^2$, if $d > 0$, $a \neq 0$ then af is positive for any x and y not both zero. Therefore, f represents numbers of only one sign, that of a. If $d < 0$, f represents both positive and negative numbers.

For binary forms with a positive determinant, the number of classes of integral, positive-definite, binary quadratic forms with a given determinant is finite.

An automorph of f is a unimodular transformation carrying f into itself. For example, $x = -x_1$, $y = -y_1$.

In the above, a, b, c may denote any real numbers. Now consider f integral; i.e., a, b, c, integers. The greatest common divisor (*g.c.d.*) of a, b, c is called the divisor of f. If the *g.c.d.* is 1, f is called primitive.

All the automorphs of the primitive form $ax^2 + bxy + cy^2$ of discriminant D are given by $x = \frac{1}{2}(t - bu)X - cuY$, $y = auX + \frac{1}{2}(t + bu)Y$ where t, u range over all the integral solutions of the so-called Pell's equation, $t^2 - Du^2 = 4$.

It can be shown that if D is a positive nonsquare integer, all integral solutions of Pell's equation are given by $\frac{1}{2}(t + u\sqrt{D}) = \pm[\frac{1}{2}(T \pm U\sqrt{D})]^k$, $k = 0, 1, 2, \ldots$, and T and U are the least solution in positive integers.

Fermat stated and Euler proved that every prime which is congruent to 1 mod 4 is the sum of two squares. For such a prime number, -1 is a quadratic residue. Lagrange found that every natural number is the sum of at most 4 squares.

Algebraic Number Theory. The notions of primality, integers, divisibility, etc., which originated in arithmetic have been extended to systems other than ordinary integers. The ordinary integers are referred to as rational integers to distinguish them from integers in fields other than the real number field. Gauss defined a complex number as $a = a_0 + a_1 i$, where a_0 and a_1 are rational and $i^2 = -1$. The conjugate of a denoted by a^* is given by $a^* = a_0 - a_1 i$.

The product aa^{**} is rational and is called the norm of a denoted by $N(a)$ which is equal to $a_0^2 + a_1^2$. If $a = bc$, then $a^* = b^*c^*$ and $aa^{**} = (bb^*) \cdot (cc^*)$ which shows that the norm of b is a factor of the norm of a. As with ordinary integers $b|a$ if $a = bc$ in complex integers. A complex integer is a number of the form $a = a_0 + a_1 i$ where a_0 and a_1 are rational integers.

The only complex numbers which divide all complex integers are those of norm 1, ± 1 and $\pm i$. If $b|a$ the associates ub of b, where u is any unit, are also factors of a. If the complex integer p has no factors other than associates and units, p is called a prime.

The complex number $p = p_0 + p_1 i$ is a complex prime if $N(p)$ is a rational prime such that $N(p) \equiv 1(\text{mod } 4)$, $p = \pm(1 \pm i)$, or if p is the associate of a rational prime which is congruent to 3 modulo 4. One can prove that $p|ab$ only if $p|a$ or $p|b$ and thus deduce the fundamental theorem of arithmetic for complex integers. Virtually all theorems in the theory of ordinary numbers have analogues for complex numbers.

An algebraic number field $R(\theta)$ of degree n is generated by the root θ of an algebraic equation $a_0x^u + a_1x^{u-1} + \cdots a_n = 0$ where all a_i are rational and $a_0 \neq 0$, is an algebraic integer if all a_i are rational integers and $a_0 = 1$. The algebraic integers in an algebraic number field form an integral domain. Consider numbers of the form $a = a_0 + a_1\theta$, where $\theta^2 = -5$. The conjugate a^* and the norm of a are defined in a manner analogous to the complex numbers. The number 21 can be expressed as $21 = (1 + 2\sqrt{-5})(1 - 2\sqrt{-5})$ or as $21 = (4 + \sqrt{-5})(4 - \sqrt{-5}\,)$, where the factors on the right hand side of each equation are prime in this system. This example demonstrates that the fundamental theorem of arithmetic (uniqueness of factorization) does not always hold true for such algebraic number fields.

Uniqueness of factorization can be restored by the introduction of ideals. An ideal a in the algebraic number field $K = R(\theta)$ is a set of integers of K such that: (1) if α and β are in a, then $(\alpha + \beta)$ is in a; (2) if α is in a and ξ is any algebraic integer in K than $\alpha\xi$ is in a. A number α is replaced by its principal ideal (α) which consists of all numbers $\alpha\xi$, where ξ runs through all integers of K.

Ideals, which were invented for use in number theory, are now used extensively in higher algebra. Other generalizations of numbers occur in the theory of linear algebras. The algebra of quaternions is an example. See also **Quaternion.**

Analytic Number Theory. This involves the use of the methods of calculus and function theory to study the properties of the integers. The most famous problem is to determine the number $\pi(x)$ of prime numbers up to X. Gauss and Legendre suggested asymptotic formulas for $\pi(x)$. The simplest formula is $\pi(x) \sim (x/\log x)$. The symbology $f(x) \sim g(x)$ indicates that $f(x)$ is asymptotic to $g(x)$. Two functions, $f(x)$ and $g(x)$, are said to be asymptotic when the ratio $f(x)/g(x)$ tends to one as a limit as x tends to infinity.

This formula was not proved until 1896. The methods involved the use of the Riemann-Zeta Function $\zeta(s)$ defined by the series $\zeta(x) = \Sigma^\infty 1/n^s$. Where s is a complex variable $s = \sigma + it$. The series is convergent only for $\sigma > 1$ but $\zeta(s)$ can be defined for $\sigma \leq 1$ by using analytic continuation and is found to be a function with only a simple pole of residue 1 at $s = 1$. By using the theorem of unique prime number factorization, it can be shown that $\zeta(x) = \Pi_p\, 1/(1 - p^{-s})$, where p runs over all prime numbers, $\zeta(s)$ has no zeros for $\sigma = 1$. Riemann knew that $\zeta(s) = 0$ for s equal any negative even integer and that all the remaining zeros satisfied $0 < \sigma < 1$. Riemann conjectured that all the nontrivial zeros of $\zeta(s)$ lie on the line $\sigma = \frac{1}{2}$. This conjecture was contained in a memoir of 1859 and remains unproved.

P. G. L. Dirichlet proved that there exist infinitely many prime numbers which satisfy $p \equiv a(\text{mod } m)$ where a and m are given coprime numbers. Dirichlet defined a more general set of functions $L(s, x) = \Sigma_{n=1}^\infty \mathcal{X}(n)/n^s$ which are known as Dirichlet L-functions. Where $s = \sigma + it$ as before and $\mathcal{X}(n)$ is a character of the multiplicative group of residue classes modulo m and coprime to m. Therefore, $\mathcal{X}$ satisfies the equations $\mathcal{X}(a_1) = \mathcal{X}(a_2)$ for $a_1 \equiv a_2(\text{mod } m)$ and $\mathcal{X}(a)\mathcal{X}(b) = \mathcal{X}(ab)$. $\mathcal{X}(n)$ is zero if n and m contain a common factor greater than 1. $\mathcal{X}(n)$ can have values ± 1, 0, or roots of unity.

The conjecture that there are no zeros of the L-functions with real part $\sigma > \frac{1}{2}$ is known as the Generalized Riemann Hypothesis.

Additive Number Theory. Srinivasa Ramanujan was a protegé of G. H. Hardy. Hardy once visited Ramanujan in a hospital in England. Hardy remarked that the taxi in which he came had the rather uninteresting number 1729 for its license number. Ramanujan immediately remarked that 1729 was interesting because it was the smallest positive integer that could be expressed as the sum of two cubes in two different ways, namely $1729 = 12^3 + 1^3 = 10^3 + 9^3$. This is an example of an additive number theory result. Partitions and polygonal numbers which occur in additive number theory are discussed below.

A partition of n is a decomposition of the number n into additive parts. Repetitions in the additive parts are allowed and the order is irrelevant. Let $p(n)$ denote the number of partitions of n. Euler gave

$$1 + \sum_{n=1}^{\infty} p(n)x^n = \prod_{k=1}^{\infty} (1 - x^k)^{-1}$$

as a generating function for $p(n)$. Generating functions can be constructed for partitions that are subject to various restrictions. Euler observed that if $E(n)$ is the number of partitions of n into an even number of unequal parts and $U(n)$ is the number of partitions into an odd number of unequal parts, then $E(n) - U(n) = 0$ unless n is of the form $(\frac{1}{2})m(3m \pm 1)$, when $E(n) - U(n) = (-1)^m$.

As asymptotic form for $p(n)$ was developed by Hardy and Ramanujan, the first term of which is

Rademacher gave an exact expression for $p(n)$ as the sum of a convergent infinite series. Ramanujan observed that $p(n)$ possesses various congruence properties such as

$$p(n) \sim \frac{1}{4n\sqrt{3}} \exp\left(\pi \frac{2n}{\sqrt{3}}\right).$$

$$p(5n + 4) \equiv 0(\text{mod } 5)$$

and

$$p(7n + 5) \equiv 0(\text{mod } 7).$$

A polygonal number of order m is given by $x + \frac{1}{2}(m - 2)(x^2 - x)$, where $x = 0, 1, 2, \ldots$. The value of m is equal to the number of sides of a polygon; i.e., $m = 3$ gives triangular numbers, $m = 4$ gives square numbers, etc. Fermat stated that every positive integer is the sum of m polygonal numbers of order m. This theorem was proved by Legendre for $m = 3$, Lagrange for $m = 4$, and by Cauchy for the remaining values of m.

Polygonal numbers of order 4 are squares of integers. Fermat's theorem says that every positive integer can be expressed as the sum of one, two, three, or four nonzero squares. Waring stated an extension of this theorem to higher powers. Every positive integer can be expressed as the sum of at most $g(k)$ of kth powers. Another function $G(k)$ which gives the least number of kth powers required to represent all but a finite number of positive integers is of even greater interest. Clearly, $G(k) \leq g(k)$. Lagrange's result indicates that $G(2) = g(2) = 4$. Other results include $g(3) = 9$, $G(3) \leq 7$, $19 \leq g(4) \leq 35$, $G(4) = 16$, $37 \leq g(5) \leq 54$, and $G(5) \leq 23$. Vinogradow proved that $G(k) \leq k(3 \log k + 10)$.

Vinogradow also proved that every sufficiently large odd number is the sum of three odd primes.

Diophantine Approximations. Every number which is processed by a computer or written in decimal form with a finite number of digits is a rational number. Therefore, there is some practical interest in the errors introduced by approximating irrational numbers by rational fractions. By using continued fractions, it can be shown that if θ is any irrational number, then there are infinitely many fractions p/q such that $|\theta - p/q| < 1/q^2$. A continued fraction is an expression of the form

$$a_1 + \cfrac{b_1}{a_2 + \cfrac{b_2}{a_3 + \cdots}}$$

which is abbreviated $a_1 + (b_1/a_2) + (b_2/a_3) + \ldots$. If all b_i's are equal to 1, the continued fraction is called a simple continued fraction. The simple continued fraction

$$a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}$$

is sometimes abbreviated as $[a_1, a_2, a_3, \ldots]$. The finite simple continued fractions

$$c_1 = [a_1] = a_1$$
$$c_2 = [a_1, a_2] = a_1 + \frac{1}{a_2}$$
$$c_3 = [a_1, a_2, a_3] = a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3}}$$

are called the convergents of the simple continued fraction $[a_1, a_2, \ldots a_k]$.

In general, it can be shown that $c_1 < c_3 < c_5 \cdots < x < \cdots c_3 > c_4 > c_2$, where x is any irrational number. If x is rational the last convergent is equal to x (in this case the continued fraction is finite).

A periodic continued fraction is a continued fraction of the form $[a_1, a_2, \ldots a_n, \overline{b_1, b_2, \ldots b_m}]$where n is a nonnegative integer and m is a positive integer. The period of the continued fraction is the sequence of repeating terms $b_1, b_2, \ldots b_m$. The length of the period is m.

A quadratic irrational is an irrational number that is a solution of $ax^2 + bx + c = 0$ where a, b, and c are integers. Every periodic continued fraction represents a quadratic irrational and every quadratic irrational can be expressed as a periodic continued fraction. It can be shown that if k is not a square then $\sqrt{k}$ can be expressed as

$$\sqrt{k} = [a_1, \overline{a_2, a_3, a_4, \ldots a_4, a_3, a_2, a_1}].$$

Liouville proved that if ξ is any real algebraic number of degree n and k is any constant, then there exist at most a finite number of rational fractions p/q such that $|p/q - \xi| < (k/q^{n+1})$. This result can be used to construct transcendental numbers. Choose a number λ with a sufficiently rapid sequence of rational approximations such that there are an infinite number of rational fractions p/q that satisfy $|p/q - \lambda| < (k/q^{n+1})$.

Donald R. Hodge, The BDM Corporation, Vienna, Virginia.

NUMERICAL CONTROL. Since the Industrial Revolution (circa 1760), there has been a continuing search for more effective manufacturing methods. As part of this unrelenting process, numerically controlled (NC) machines were developed in the early 1960s and have been and continue to be widely used. These developments occurred before the widespread use of electronic computers, and, of course, the personal computer was unknown at that time. The simplest NC machines of the 1960s were *open-loop*—that is, the two or three axes of a machine tool were directed to manipulate the machining of a part in two or three dimensions by controlling the axes of the machine. There was no feedback to the controller. See Fig. 1. However, *closed-loop* systems with feedback to the controller were introduced shortly thereafter. See Fig. 2. Numerically coded coordinate data were fed to the early machines by means of punched paper (later mylar) tape. The tapes proved to be cumbersome to prepare and very difficult to store and retrieve over long periods of time. As a consequence of these inconveniences, but limited to rather large installations, data were transmitted to the controller by means of digitized electric signals. Unfortunately, these modernized early systems were limited to sharing mainframe computers because smaller computers with the required capacity were still unavailable. Later, less expensive computers with the required data-handling capacity were introduced. These led to the tying of NC systems with CAD/CAM data centers. In what is now termed *distributed numerical control* (DNC), one computer can feed part programs to *multiple* machine tools. The advent of mini- and microcomputers, plus the ability to address multiple serial ports on a real-time basis, gained acceptance of this technology by manufacturers. Contemporary NC systems continue to cut down job lot sizes, reduce setup time, and trim machining time and the skill level of the direct labor force. It should be emphasized further that NC systems are now tied together at the factory floor level by *local area networks* and that these, in turn, can receive and feed information to plant-wide and corporate-wide networks.

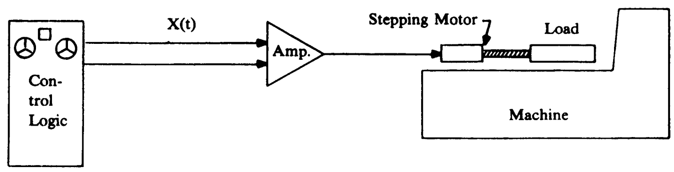


Fig. 1. Open-loop numerical control system. (*Circa early 1960s.*)

NC Compoents. In the early days of applying NC, machines were not designed with NC in mind and lacked such characteristics as stiffness, accuracy obtainable by a highly skilled worker, and often the required production speed. Today, many machines are designed specifically with NC as an objective. Much attention is paid to position measurement accuracy and by devices that have electronic outputs. Two types of positioning devices have emerged. They are available in both linear and rotary form. Linear devices possess very good accuracy because they are a direct measure of axis position, but they also are somewhat difficult to mount and protect. Rotary types are more compact, but may be more subject to wear and accuracy deterioration because they

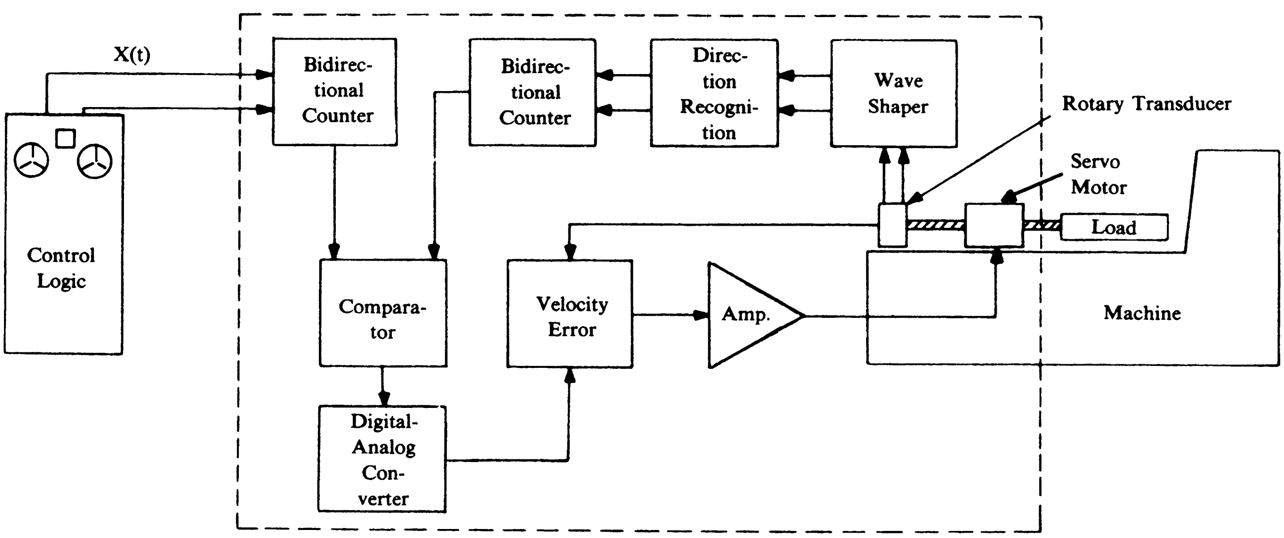


Fig. 2. Closed-loop numerical control system.

measure position by way of lead screws or rack-and-pinion methods with gearing. See also **Position and Displacement Measurement.**

Designing for NC Production

Although NC has been a major factor in achieving flexible factory automation, this does not mean that plant engineers have *carte blanche* in terms of designing parts without consideration of currently installed manufacturing methods. Indeed, for the high levels of productivity and quality needed for a business to survive, this knowledge is now more important than ever. While cells and flexible manufacturing systems add to the flexibility of a plant, they also have considerations that the designer needs to be aware of in order to make optimum use of these high-priced installations.

The size and shape of the part are among the most obvious and important criteria. Most manufacturers define their machine capabilities in terms of "cube" of workspace that can be accommodated. This does not mean that the overall part size must fit into this cube, but rather that machining operations must be limited to it. Workpiece weight is another consideration for table-type machines. Where floor-type machines are available, these problems are less critical. Cells and flexible manufacturing systems introduce the additional consideration of the material handling equipment. Size and weight are particularly important where parts are to be moved around automatically, and these limitations must also be known by the designer.

The number and type of cutting tools required to machine a workpiece must also be in the forefront of the designer's mind. The capacity of automatic tool-change equipment ranges from about ten to one-hundred tools. Significant increases in manufacturing efficiency can be achieved if these do not require frequent change. In particular, tool-change time and thus machine-cycle time is reduced and errors caused by incorrect tools being put into tool-changer sockets can be eliminated. These factors are particularly important in cells and flexible manufacturing systems where several machines operating on several parts or families of parts are simultaneously in production. The major saving in tool quantity is still probably to be achieved in the area of tapped holes that often require three tools per thread type. As a byproduct of saving tool magazine space, tool-change time is also reduced, thereby cutting down on total machining time.

Another problem made more critical by the advent of multiple machine cells and flexible manufacturing systems is that of holding the workpieces for machining. Not only must the part be held while being cut, but it also must be transported between machines and the system or cell load and unload stations. Where designers can help is by keeping required machining operations to as few sides of the part as possible and by providing surfaces or other features to locate and clamp on quickly and effectively. In most systems, all load and unload as well as refixturing operations must be done at special stations and not on the machine tools themselves. Thus, each refixturing requires the part leaving a machine, traveling to a special station, and then traveling back again to continue being machined. This involves significant time as well as making the material transporter unavailable for other operations and can have an important detrimental effect on the efficiency of the system as a whole.

The designer must also know the rotary axis capability of the manufacturing plant. Many rotary tables can only position in 15-degree steps, and any subdivisions beyond this angle can result in severe difficulties for the shop floor. The cost increment between tables with indexing as opposed to full positioning capability to any angle can be very large, and thus designs that call for this requirement in order to be manufactured may be very expensive.

The designer must also keep in mind the part programmer and the capabilities available to simplify the task of producing effective NC tapes. Features such as origin shift, subroutines, probing, etc., previously mentioned are targeted at making programming easier in cer-

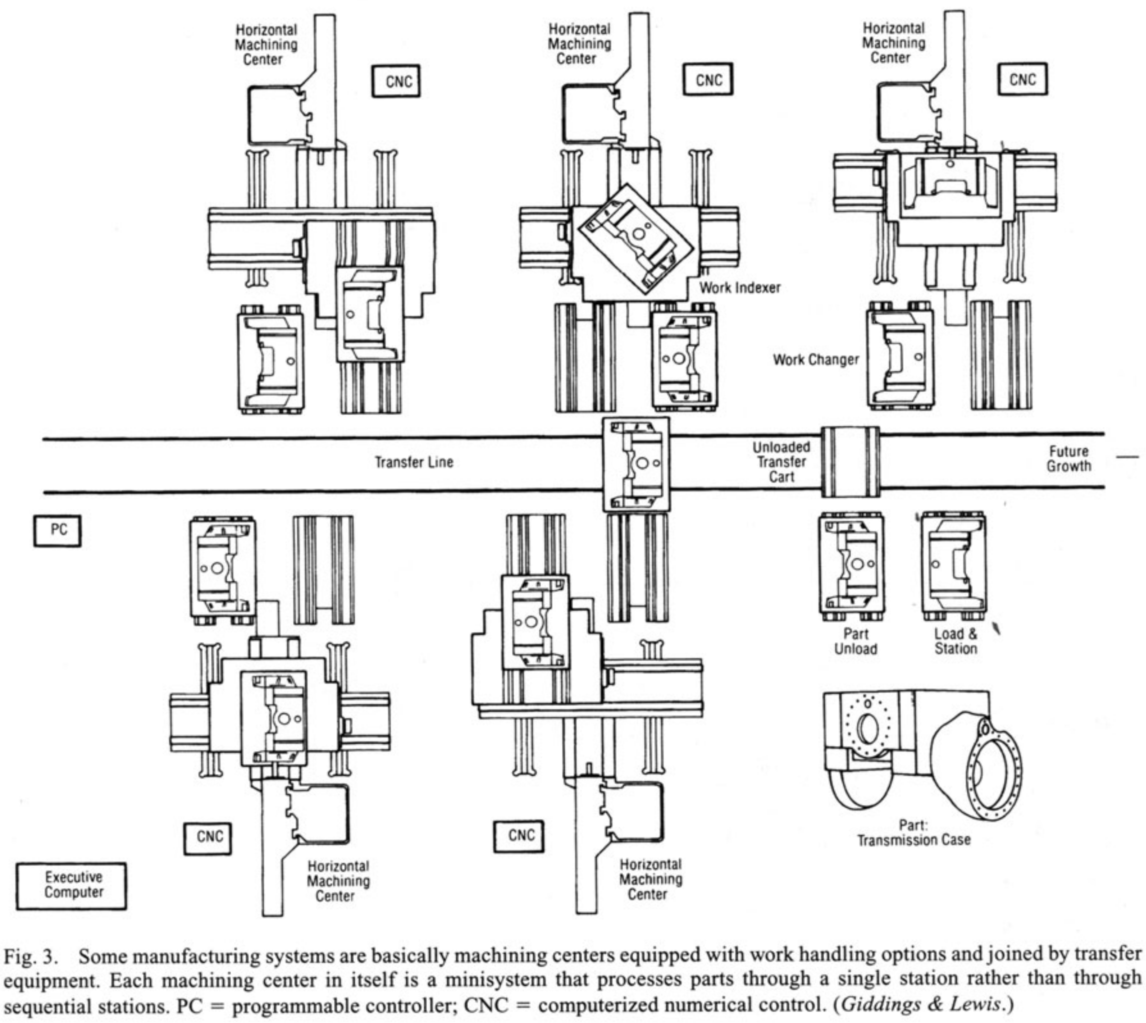


Fig. 3. Some manufacturing systems are basically machining centers equipped with work handling options and joined by transfer equipment. Each machining center in itself is a minisystem that processes parts through a single station rather than through sequential stations. PC = programmable controller; CNC = computerized numerical control. (*Giddings & Lewis.*)

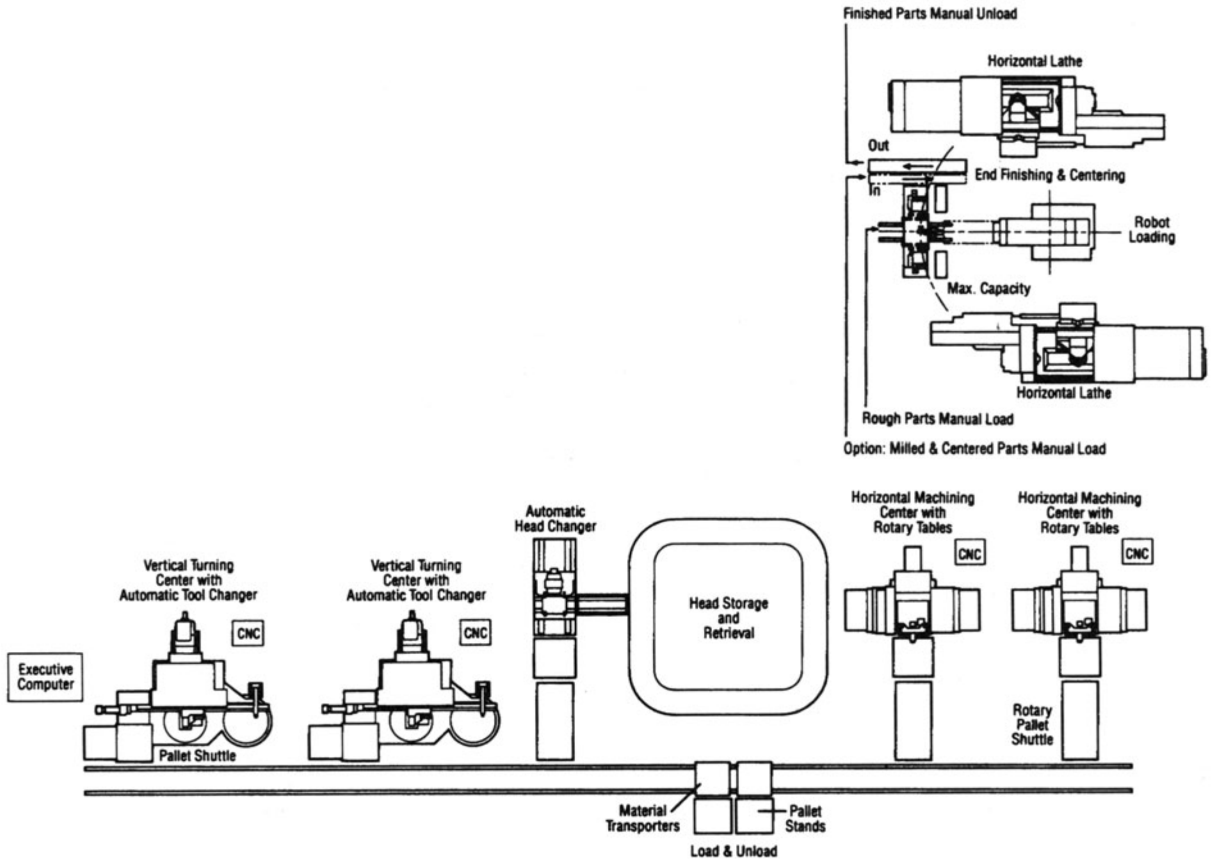


Fig. 4. Mid-volume manufacturing is frequently best served with general-purpose equipment and controls arranged as production modules. Thus, familiar and proven parts and components can be programmed by a master computer that has been preprogrammed for any number of parts. (*Giddings & Lewis.*)

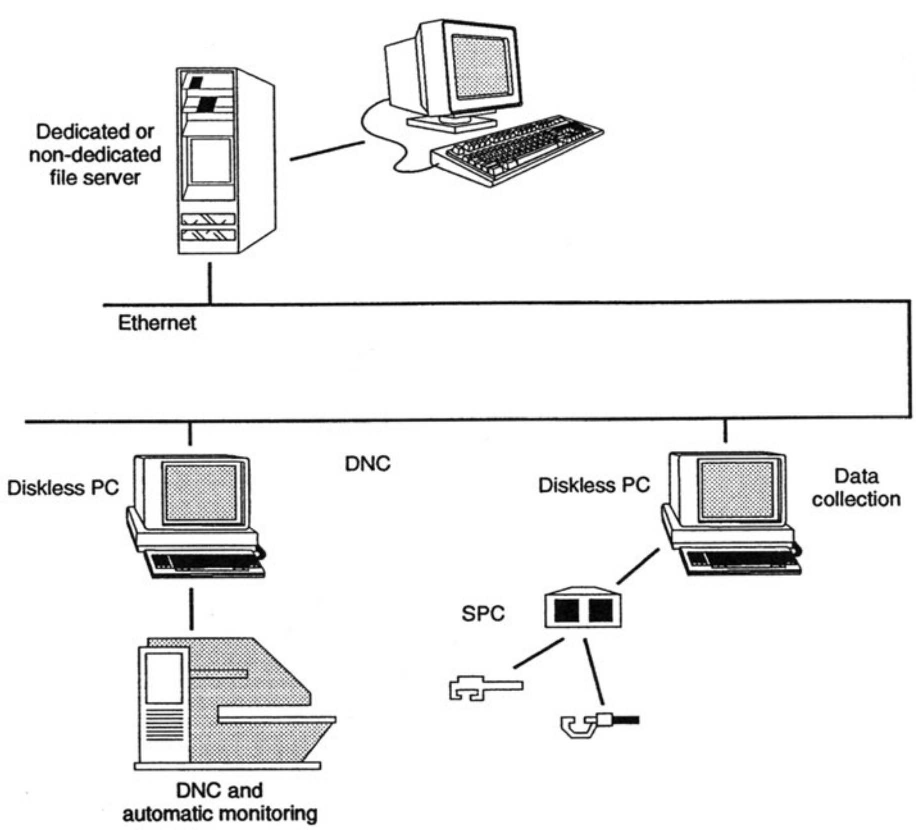


Fig. 5. Entry-level network for direct numerical control (DNC).

Basic system: (1) Entry-level standard network, (2) dedicated or nondedicated file server, (3) based on Intel 286 microprocessor, (4) supports up to five active users, (5) controls application information over network, (6) applications can be integrated in seamless environment through factory network control system.

Features: (1) High performance—files are transferred at 10 Mbs, (2) easy-to-use menu-driven utilities put the network supervisor in control, (3) security—allows supervisor to restrict network access, (4) single-source database for complete control of all files, (5) cost-effective network uses diskless PCs, (6) virtually unlimited expansion capabilities.

Applications: Statistical process control (SPC), (2) direct numerical control (DNC), (3) data collection, (4) view graphics and documentation, (5) automatic monitoring. (*CAD/CAM Integration, Inc.*)

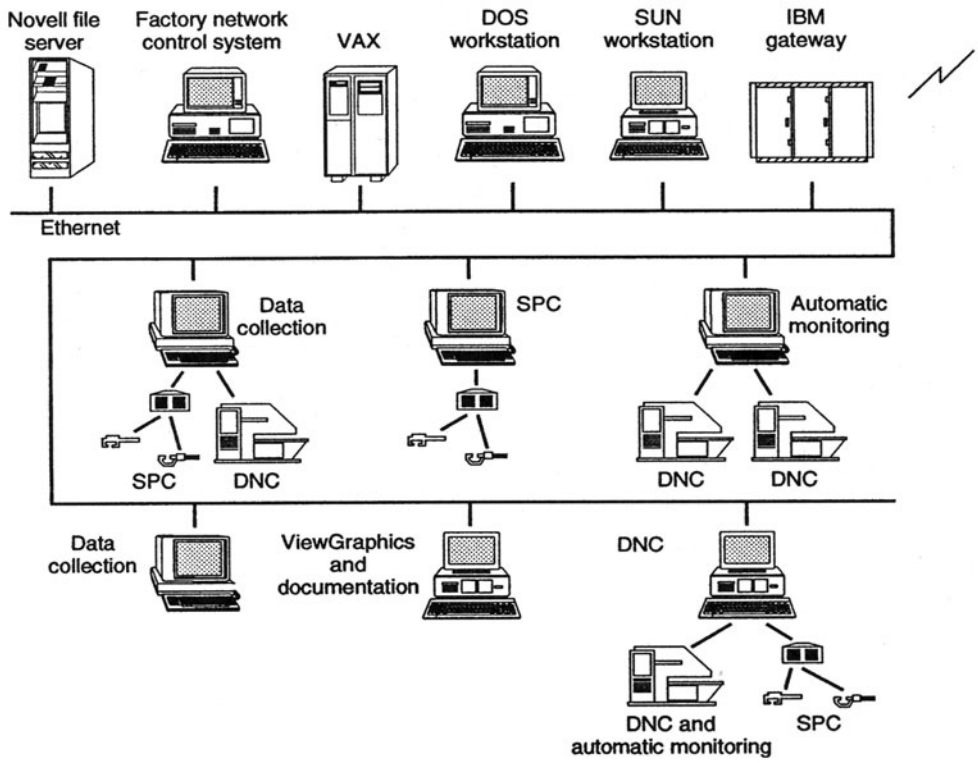


Fig. 6. Advanced NC network.

Basic system: (1) Controls application information of network, (2) applications can be integrated in seamless environment through factory network control system, (3) designed for medium to large business, (4) dedicated 386-based file server technology, (5) supports up to 100 active users on network, (6) high-performance network operating system.

Features: (1) Provides high functionality while maintaining high performance levels, (2) multiuser, multi-tasking architecture allows user to perform many operations simultaneously, (3) enhanced features for security, system reliability, and network management, (4) cell configurations decrease exposure to manufacturing down-time, (5) single-source database for complete control of files, (6) cost-effective network—uses diskless PCs, (7) simple-to-use menu-driven utilities make learning and adding applications easy.

Applications: (1) Statistical process control (SPC), (2) direct numerical control (DNC), (3) data collection, (4) view graphics and documentation, (5) automatic monitoring. (*CAD/CAM Integration, Inc.*)

tain commonly found manufacturing situations. By designing the part to take advantage of these features, the designer can enhance the success of the manufacturing process.

Numerical control can be an effective production tool, not only for the very large manufacturing complexes, but for smaller shops as well. A panorama of contemporary systems is shown by Figures 3 through 6.

NUSSELT NUMBER. The nondimensional parameter, defined as

$$N_u = \frac{Q}{\Delta T}\frac{d}{k}$$

where Q is the heat loss or heat transfer from a solid body, ΔT is the difference of temperature between the body and its surroundings, d is the scale size of the body, k is the thermal conductivity of the surrounding fluid. The Nusselt number is useful in the reduction of measurements of free and forced convective loss of heat either from the same body in different conditions or from different bodies of geometrically similar shapes. See also **Heat Transfer.**

NUTATION. In the case of a spinning object (e.g., a top or gyroscope, a particle, or an astronomical body), the inclination of the axis to the vertical will vary periodically between certain limiting angles. This motion is called nutation. It is a variation in precession. In astronomy, nutation is caused by the attracting force of the sun and the moon tending to pull the equatorial bulge of the earth into the plane of the ecliptic. The amount of this force is changing throughout the year as the declinations of the sun and the moon change. For example, twice during each year, both the sun and the moon are on the equator, and at those times, their precessional forces are zero. The principal nutation is due to the periodic change in the plane of the moon's orbit, and has a period of about 19 years. Most of the nutation effects are periodic in character.

NUTHATCH (*Aves, Passeriformes*). Small climbing birds which cling in any position to the bark of trees as they search for food. They eat both insects and seeds. The most common North American species is the white-breasted nuthatch, *Sitta carolinensis*, found throughout the United States east of the Rockies. The red-breasted nuthatch, *S. canadensis*, is more common in Canada and the mountains but migrates in winter as far as the southern states. The pigmy nuthatch, *S. pygmaea*, is a western species. Nuthatches are found on all other continents except South America, although in Africa they are confined to the north.

NUTMEG. See **Flavorings.**

NUTMEG TREE. Of the family *Myristicaceae*, the nutmeg tree grows to about 60 feet (18 meters) in height and has pointed lanceolate leaves. Nutmegs are the seeds of this tree which is native to the Molucca Islands. The trees are dioecious, pistillate and staminate flowers being borne on separate trees. Since the trees are frequently grown in cultivation, especially in favorable localities, it is necessary to plant some of both sexes to ensure cross-pollination and seed formation. The flowers are pale yellow, the fruit a dark orange-colored berry containing a single large brown seed. Surrounding the seed is a branched deep-red aril, which upon drying becomes pale brown. The seed is the nutmeg of commerce; the aril is mace. Both the seed and the aril contain an aromatic oil, but only the poor-quality seeds are used for extraction of oil. The tree begins to bear at about 6 years of age and may produce for nearly a century. The tree produces best when planted at an altitude of from 700 to 1,800 feet (213 to 549 meters) in a tropical climate. The average tree will produce about 20 pounds (9 kilograms) of nutmegs per year. An acre produces about 1,200 pounds (544 kilograms) of green nutmegs annually. From this quantity, about 150 pounds (68 kilograms) of mace are produced; and 720 pounds (327 kilograms) of dried nutmegs. See also **Flavorings.**

NUTRIENTS (Soil). See **Fertilizers.**

NUX VOMICA TREE. Of the family *Loganiaceae*, the nux vomica tree is a small-to-moderate size tree found in India, Sri Lanka, Burma, Thailand, and Australia. The trunk is often crooked, short and thick. The leaf is smooth with three to five veins and ovate. The flower is light green, small and of a tubular shape. The fruit is hard with a gelatinous pulp. The tangerine-size fruit contains from one to five disk-shaped seeds. The seeds contain powerful alkaloids, including strychnine and brucine.

NYMPH. If the wing pads of an insect are developed on the outside of its body, the insect is said to have a *simple* or *gradual metamorphosis*. The insect during this growing stage is called a *nymph*. This situation is contrasted with the case where the wing pads are developed internally during the growing stage. In this case, the growing stage is called a *larva*, and the insect is daid to have a *complete* or *complex metamorphosis*. Nymphs generally appear something like miniature adults and have the same general life style. In most cases of larva, the size, shape, locomotion, and eating habits of the larva contrast sharply with the adult form. See also **Larva.**

NYQUIST FREQUENCY. For data defined at equal time-intervals t, the frequency of a sine or cosine term with a period double the interval t. Frequencies higher than this will not be directly detectable by spectral analysis.

NYQUIST RATE (Signaling). In transmission, if the essential frequency range is limited to B cycles per second, $2B$ is the maximum number of code elements per second that can be unambiguously resolved, assuming the peak interference is less than half a quantum step. This rate is generally referred to as signaling at the Nyquist rate, and $\frac{1}{2}B$ is called the Nyquist interval.

NYSTAGMUS. See **Vision and the Eye.**

NYTRIL. See **Fibers.**